UNIVERSITY OF LONDON THESIS

Degree PhD Year 2007 Name of Author HAQ, Modhumita Afsana

The AmiR/RNA Interaction of the Amidase Operon Regulatory System of *Pseudomonas aeruginosa.*

Modhumita Afsana Haq

Thesis submitted to University College London for the degree of Doctor of Philosophy

Department of Biochemistry and Molecular Biology
UCL
Gower Street
London WC1E 6BT

June 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

I confirm that the work presented in this thesis is my own and the work of any other persons is appropriately acknowledged.

Modhumita Afsana Haq

ABSTRACT

Expression of the amidase operon of *Pseudomonas aeruginosa* is regulated by the AmiC and AmiR proteins using a version of the bacterial two-component signal transduction system. The negative regulator, AmiC is the amide ligand sensor and regulates activity of AmiR by a steric hindrance mechanism. AmiR, the response regulator functions as a transcription antitermination factor. Once released from the AmiC/AmiR complex, AmiR binds to the operon leader transcript to prevent formation of the rho-independent terminator thus allowing transcription of the entire operon. AmiR has a CheY-like N-terminal response regulator receiver domain without the conserved phosphate-acceptor residues and a long coiled-coil C-terminal domain terminating in a three-helix bundle ANTAR domain. Within this family are a number of highly conserved residues. These residues have been changed by site-directed mutagenesis and the mutants tested for antitermination activity by amidase assay. Random mutagenesis has also revealed key residues in this domain, changes which led to reduced antitermination efficacy. To define the minimal transcript length sufficient for AmiR antitermination, constructs have been made and analysed containing leader region mutations and deletions. The results show that the encoded leader ORF has no biological function and that sequences upstream of the previously defined L recognition region are not necessary for AmiR-dependent antitermination. The ANTAR domain has been expressed and purified as a cleavable, GST tagged fusion protein in *E. coli*. Dynamic light scattering data and gel filtration elution profiles of ANTAR suggests this domain is a dimer in solution. The ANTAR-RNA interaction was investigated *in vitro* using band shift assays. Radiolabelled *in vitro* transcribed *ami* leader RNA was incubated with the ANTAR protein under various buffer conditions. The mixtures were analysed by PAGE to monitor the formation of protein–RNA complexes. Amidase assays were also performed with the ANTAR domain for analysis of antitermination activity *in vivo*. These results show that the ANTAR domain alone is not sufficient for antitermination *in vivo* or binding to the RNA *in vitro*.

ACKNOWLEDGEMENTS

I would like to thank Rob Drew for his guidance and supervision throughout this project. His extensive background and wealth of knowledge in the field of bacterial transcription has led to some exciting advances in understanding the elegant antitermination mechanism of the amidase operon.

I would also like to thank Leon D'Cruz for the MALDI-TOF analysis of ANTAR, and help with the Dynamic Light Scattering experiments. I am grateful to Finn Werner for teaching me how to carry out those tricky RNA-binding assays, for the gift of the F/E protein, and for allowing us all access to a seemingly endless supply of wonderful music.

Thanks must go out to John Ward - and for so much. I am grateful for his generosity over the years in letting me use the equipment and facilities in his lab and I am especially grateful for being made to feel part of the lab. I would also like to thank the members of the 'John Ward lab' in particular, Ursula and Martin for all of the memorable times.

Thanks to my family for their wealth of support, encouragement and unconditional love. A very special thank-you to my little Rakha! I would also like to thank Nick, for being a great partner and life-coach to me and so much more.

Finally I would like to thank the BBSRC for funding this project.

ABBREVIATIONS

A	adenine
A	Alanine [Ala]
ADP	adenosine diphosphate
Amp	ampicillin
ANTAR	AmiR NasR Transcription Antitermination Regulators
ATP	adenosine triphosphate
Bp	base pair
C	cytosine
cAMP	cyclic adenosine monophosphate
CAP	catabolite activator protein
CAT	co-antiterminator
CCR	carbon catabolite repression
CDS	coding sequence
Cm	chloramphenicol
CRC	catabolite repression control
CRP	cAMP receptor protein
CTD	C Terminal Domain
dATP	adenosine deoxyribonucleoside triphosphate
dCTP	cytosine deoxyribonucleoside triphosphate
dGTP	guanosine deoxyribonucleoside triphosphate
dNTP	equimolar mixture of dATP, dCTP, dGTP dTTP
dTTP	thymine deoxyribonucleoside triphosphate
DEPC	diethylpyrocarbonate
dH_2O	distilled water
DLS	dynamic light scattering
DNA	deoxyribonucleic acid

E	Glutamate
EDTA	ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
FNR	Fumarate Nitrate Reduction regulator
G	guanine
GDP	guanosine diphosphate
GTP	guanosine triphosphate
GST	glutathione S-transferase
H / His	Histidine
IPTG	isopropyl-beta-D-thiogalactopyranoside
MALDI	Matrix Assisted Laser Desorption Ionisation
MCS	multiple cloning site
Mbp	million base pair
Mw	molecular weight
NMR	nuclear magnetic resonance
nt	nucleotide
NTD	N Terminal Domain
Nut	N utilisation
OD	optical density
OMP	outer membrane protein
ORF	open reading frame
ori	origin of replication
PEP	phosphoenolpyruvate
PRD	phosphotransferase regulation domain
PTS	phosphotransferase system
R	Arginine [Arg]
RAT	ribonucleic antiterminator
RNA	ribonucleic acid
RNAP	RNA Polymerase

P	Proline [Pro]
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCS	photon correlation spectroscopy
Q	Glutamine [Gln]
SA	specific activity
SD	standard deviation
S/D	Shine-Dalgarno
SDM	site directed mutagenesis
SDS	sodium dodecyl sulphate
Sm	streptomycin
TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
TCA	tricarboxylic acid
TEMED	N,N,N',N'-Tetramethylethylenediamine
TG	Tris-Glycine
Tir	Translation Initiation Regions
U	uracil
UTR	untranslated region
V	Valine
WT	Wild type
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-ga*lac*topyranoside

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Biology of Pseudomonas

Pseudomonads are Gram-negative, non-sporulating, aerobic bacteria, found in soil, decaying organic matter, vegetation and water. There is much diversity found amongst the *Pseudomonas* genus and this is partly due to the fact that for many years, since the term *Pseudomonas* was first coined in 1895, organisms that did not fit into any other taxonomic group were placed into this one. More recently, a new method of taxonomy has been developed based on classing *Pseudomonas* species by rRNA gene homology, which has led to the sub-division of Pseudomonads into five distinct groups [Palleroni 1992]. *Pseudomonas aeruginosa* belongs to rRNA gene homology group I, and bacteria found in this group are commonly known as 'true pseudomonads'.

Pseudomonads have medical, industrial, scientific and environmental importance. Medically, they are studied for the wide range of opportunistic infections for which they are responsible. *P. aeruginosa* is responsible for bacteraemia in burn victims, urinary tract infection in catheterised patients, nosocomial blood infections [Marra *et al.*, 2006; Kramer *et al.*, 2006] and in patients suffering from cystic fibrosis, a chronic localised lung infection, which is often fatal [Doring *et al.*, 2006]. The pathogenesis of *P. aeruginosa* is aided by their apparent intrinsic resistance to a wide range of antibiotics and some disinfectants. To date, *P. aeruginosa* has approximately 6,000 genes within its genome [Stover *et al.*, 2000], of which 150 code for outer membrane proteins [OMP], which is disproportionately large compared to other bacterial genomes. Many of these OMPs are involved in drug efflux systems, and confer some resistance to antimicrobial agents. In addition to this, Pseudomonads are also able to grow in biofilms [Costerton *et al.*, 1999],

which may contribute to the impermeability exhibited by the cells to therapeutic concentrations of antibiotics.

The nutritional versatility of Pseudomonads and ability to thrive in diverse, ecological environments is reflected in their large genome size, and at 6.3 million base pair [Mbp], strain PAO1, was the largest bacterial genome to be sequenced when published [Stover *et al.*, 2000]. The complete genome sequence of *P. aeruginosa* strain POA1, the wild type strain most commonly used for genetic and biochemical studies in *Pseudomonas*, was determined in a collaboration with the Cystic Fibrosis Foundation, the PathoGenesis Corporation and the University of Washington Genome Centre [Stover *et al.*, 2000], [http://www.pseudomonas.com/].

1.1.1 Discovery of an Inducible Amidase Produced by P. aeruginosa

Pseudomonads are also studied for their ability to utilise a wide range of organic compounds such as carbon and nitrogen sources and these include carbohydrates, aromatic compounds, hydrocarbons, aliphatic acids, amines and amides. In 1959 a strain of *P. aeruginosa* was isolated that was able to utilise acetamide or propionamide as sole carbon and nitrogen sources [Kelly and Clarke 1962]. Both of the amides were able to induce synthesis of a chromosomally encoded amidase, capable of hydrolysing aliphatic amides to their corresponding organic acid and ammonia. The amidase genes were subsequently cloned from a constitutive mutant [PAC433] [Drew *et al.*, 1980] and the wild type [PAC1] [Wilson and Drew, 1991] and a 5.3kb *Hin*dIII/*Sal*I restriction fragment containing the genes was sub-cloned into pBR322 to generate plasmids pJB950 and pAS20 respectively. Sequencing of the entire 5.3kb fragment, identified five genes, which were found to form the entire amidase operon; *amiEBCRS* [chapter 1.4]. The *amiC* and *amiR* gene products form an amide sensitive, two-component signal transduction system, regulating expression of the amidase operon by a novel antitermination mechanism.

This study investigates the mechanism by which AmiR antiterminates transcription of the amidase operon.

1.2 Gene Expression in Bacteria

The first evidence to emerge that genes specify the amino-acid sequence of proteins came when Ingram [1957] linked a defect in the gene that caused sickle-cell anaemia with a single amino-acid change in the β-globulin component of the haemoglobin protein. The question that arose was how the information contained in the DNA molecule was transferred into protein. It was apparent that DNA could not be directly involved because, what was thought correct at the time, enuclear cells could still synthesize protein. One theory was that a transient intermediate molecule must carry the message from the DNA in the nucleus to the cytoplasm, where proteins were made. An obvious candidate for this messenger was RNA, as it was found both inside the nucleus and in the cytoplasm, and appeared to accumulate where proteins were being produced. Several other laboratories had discovered an enzyme that had the ability to add nucleotides onto the end of an RNA molecule. However, the enzyme responsible was unable to produce full-length RNA molecules.

This was investigated by Weiss and Gladstone [1959], who designed an experiment involving the placement of an radioactive phosphate at the α-position of nucleoside triphosphates [NTP], which would ultimately become incorporated into the sugar–phosphate backbone of nucleic acids. Incubation of labelled Cytidine 5'-triphosphate [CTP] with the liver homogenate led to some of the label being incorporated into the RNA fraction. All four of the NTPs were required for the reaction and was the first step in showing that DNA provided the template for the synthesis of RNA. Digesting the labelled product back to mononucleotides showed that each of the nucleotides carried a portion of the label, showing that the CTP was

incorporated next to each of the other nucleotides as part of a growing chain, rather than just being added on to the end. This was the first description of an RNA polymerase.

1.2.1 Induction of Gene Expression

Bacteria have developed global and specific control mechanisms that enable them to sense the nutritional conditions of their environment and to adjust their metabolic pathways for carbohydrate utilisation by regulating gene expression accordingly. In a seminal paper in 1961, Jacob and Monod proposed a general model for bacterial gene regulation from their genetic study of the *Escherichia coli* lactose operon. The study of bacterial transcription initiation in *E. coli* as a model, has since served as a basis for understanding transcriptional control throughout all kingdoms of life.

In addition to specific control of operon expression in *E. coli*, inducible catabolic operons are 'globally' regulated by a catabolite activator protein [CAP], also known as cAMP receptor protein [CRP], containing bound cyclic AMP [cAMP]. Another global regulator of gene expression in *E.coli* is the Fumarate and Nitrate Reduction regulator protein, [FNR], which activates genes in response to oxygen starvation. Oxygen is detected via an iron sulphur cluster located in its NTD. Under oxygen starved conditions it is able to bind target promoters and regulate their transcription. FNR is thought to regulate approximately 100 operons, although it may indirectly affect up 1000 genes in *E.coli* [Grainger *et al.*, 2007]. CRP and FNR are global regulators that activate transcription at a large number of different promoters and are two examples of seven global transcription factors. Here, CRP regulation will be examined more closely.

1.2.2 Carbon Catabolite Repression and CRP

Carbon catabolite repression is a means by which glucose levels dictate whether or not operons for the utilisation of other sugars are induced when those alternate sugars are present in the medium. In 1961, Monod discovered this phenomenon when he showed that *E. coli*, in the presence of a mixture of two sugars, one of the sugars is metabolised first, followed by the second sugar and described this phenomenon as 'diauxie'.

Expression of the *lac* operon is inhibited in the presence of glucose and its catabolites, but increased levels of cyclic AMP [cAMP] reverse this inhibitory effect. Zubay and others [1970], proposed that cAMP binds to a protein factor, CRP, causing it to bind to a DNA site in the *lac* promoter and stimulate transcription initiation. Using the first cell-free system for transcription and translation, they showed directly that CRP plus cAMP stimulated the expression of the *lac* operon. Thus, in *E. coli*, catabolite repression operates as a positive control system, and is linked to the fact that in cells grown in the presence of glucose; there are low intracellular levels of cAMP. CRP is one of seven global transcription factors, which is estimated to regulate expression of more half of the genes in *E.coli*.

Synthesis of cAMP, which is catalysed by adenylate cyclase, is regulated by a phosphoenolpyruvate [PEP]-protein phosphotransferase system [PTS] [Ye and Saier, 1996]. This mechanism involves a central regulatory protein of the PTS, glucose specific enzyme [IIAGlc]. When glucose is present in the bacterial growth medium, IIAGlc becomes dephosphorylated as the phosphoryl groups are transferred to incoming glucose molecules. In the absence of glucose, phosphorylated IIAGlc binds to adenylate cyclase thereby increasing cAMP synthesis. cAMP binds to dimeric CRP molecules which then bind to the *lac*

promoter region which subsequently causes transcriptional activation of the *lac* operon.

Expression of CRP, encoded by the *crp* gene, is regulated by the cAMP/CRP complex [Cossart and Gicquel-Sanzey, 1985]. The cAMP/CAP complex both inhibits and activates transcription of *crp*. Thus in the presence of glucose, [low cAMP], CRP levels decrease and in the absence of glucose, [high cAMP], CRP levels increase. It appears that this mechanism enhances catabolite repression in the presence of glucose by causing deprivation of CRP mediated activation [Ye and Saier, 1996]. CRP is a factor dependent global activator protein with two functional domains; one of which acts as a sensor [cAMP binding domain] and the second is a DNA binding domain. CRP binds to DNA non-specifically but the cAMP/CRP complex assumes a change in conformation, which enables it to bind to a consensus sequence upstream of a variety of catabolic operon promoters to increase transcription rates. DNA bound CRP makes contact with the α subunit of RNA polymerase and the precise interaction depends on whether the CRP binding site is near the promoter or whether the activator binding site for CRP is more distant. Additionally, cAMP/CRP binding causes severe DNA bending to occur near to the centre of the CRP binding site [Zinkel and Crothers, 1991]. It is likely that this DNA bending facilitates protein-protein interactions in the transcriptional initiation complex [Ye and Saier, 1996].

The molecular mechanisms of catabolite repression control have been extensively studied in enteric bacteria, where glucose is the preferred carbon source. *P. aeruginosa* uses tricarboxylic acid [TCA] cycle intermediates [which include organic acids such as succinate] as carbon sources preferentially over carbohydrates such as glucose or mannitol, and appear to do so in a cyclic-AMP-independent manner. Phillips and Mulfinger [1981] showed that in contrast to *E. coli*, *Pseudomonas* levels of adenylate cyclase activity, cAMP phosphodiesterase

activity, and cAMP pools do not fluctuate with carbon source, nor does the addition of cAMP relieve repression of catabolite responsive pathways. In addition, only one PTS system, fructose, has been identified in *Pseudomonas* [Durham and Phibbs, 1982], suggesting that PTS components are not involved in catabolite repression control in Pseudomonads. The only protein thus far shown to be involved in catabolite repression as a global carbon metabolism regulator in *Pseudomonas* is the catabolite repression control [CRC] protein of *P. aeruginosa*. Crc is involved in the regulation of the expression of branched-chain keto acid dehydrogenase, glucose-6-phosphate dehydrogenase, and amidase [Hester *et al.*, 2000], but a specific function has not been assigned for this protein.

In addition to the global regulatory strategies that bacteria employ to regulate induction of gene expression, a more specific type of control can take place at any stage, both prior to and post the initiation of transcription, these stages are also targets for regulation. Central to these events is the enzyme DNA dependent RNA polymerase.

1.3 RNA Polymerase and the Sigma Factors

In all cellular organisms, transcription from DNA to RNA is synthesized by an intricate molecular machine. DNA-dependent-RNA-Polymerases are the enzymes responsible for the transcription of RNA. RNA polymerase [RNAP], is a central component to the formation of the transcription initiation complex and all subsequent stages of transcription. Repeated attempts to obtain three-dimensional crystal structures of the well-studied *E. coli* RNAP have failed and these efforts have been abandoned in favour of the high-resolution structures of RNAP obtained from thermophilic bacteria. Darst and co-workers solved the structure of the bacterial RNAP core from *Thermus aquaticus* at 3.3 Å [Zhang *et al.*, 1999], and Kornberg and others determined the structure of yeast RNAP II to 3.5 Å [Cramer *et*

al., 2000]. These structures revealed a crab-claw-shaped molecule with two pincers and a central cleft. The active centre was located on the floor of the cleft, where three highly conserved aspartic-acid residues chelate a Mg^{2+} ion. These structures were solved in the absence of templates, substrates or products, and had to be modelled based on the low-resolution structures of RNAP [Figure 1.1]. Both groups proposed that the DNA template binds at the active-centre in the cleft, following a sharply bent path, and that nucleoside 5'-triphosphate substrates enter the active-centre through one tunnel, and the RNA product exits the active-centre through another.

The core enzymes of the multisubunit RNAPs from bacteria, archaea and eukaryotes show striking similarities to each other and the amino acid sequences of the two largest subunits of cellular RNAP are conserved from bacteria to man [Allison *et al.*, 1998]. Indeed, sequence conservation indicates structural and functional homologies, however, the structures of *prokaryotic* and *eukaryotic* enzymes show more conservation than could be predicted from sequence alone [Kuznedelov *et al.*, 2006].

Transcription of eukaryotic DNA requires three different RNA polymerases: RNA polymerase I [RNAP I], RNA polymerase II [RNAP II], and RNA polymerase III [RNAP III]. Each works on different sets of genes and there is no interchangeability. RNAP II is responsible for transcribing protein coding genes and most small nuclear RNA [snRNA] genes. RNAP II is the most extensively studied and is most similar to bacterial RNAP. In bacteria, RNAP contains a catalytic core with a subunit composition $\alpha 2\beta\beta'\omega$, which associates with a Sigma [σ] subunit to form RNAP holoenzyme [Table 1.1.] [Polyakov *et al.*, 1995]. The core enzyme, in the absence of σ, is fully active in RNA polymerisation, but is unable to initiate transcription. The sigma subunit recognises promoter regions and initiates transcription. Sigma contributes to specific initiation by decreasing the

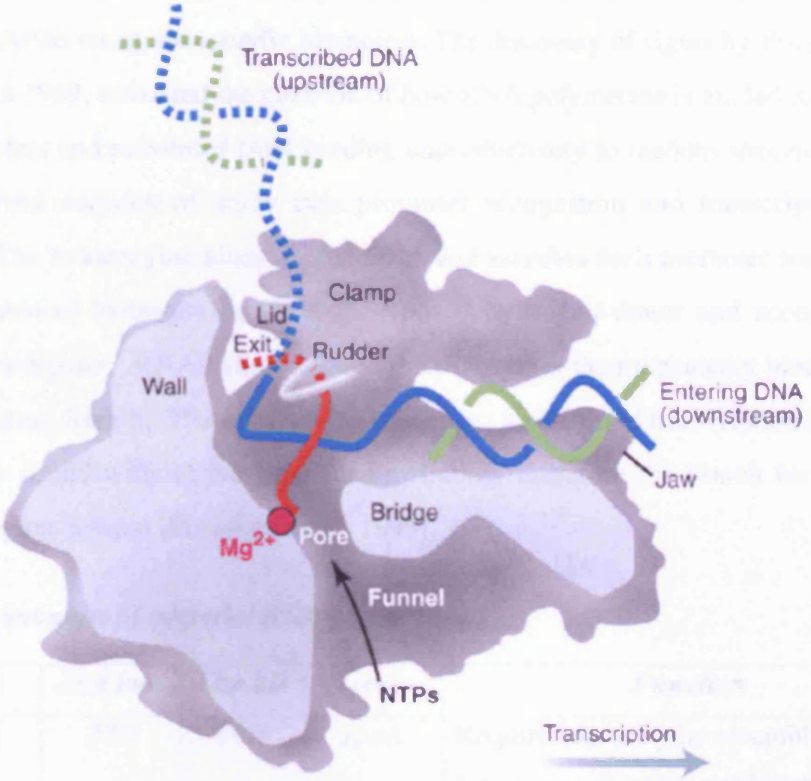


Figure 1.1. The RNA Polymerase II Elongation Complex

Sliced side view of the RNA polymerase II complex, showing the paths of the nucleic acids and the locations of some of the functional features. The sliced surfaces of the protein in the front, are shaded light grey and the remainder is in dark grey. The template strand coding for RNA is in blue, the non-template strand is in green, and the RNA in the DNA-RNA hybrid in the active centre is in red. The 3' growing end of the RNA is located adjacent to an active site Mg^{2+} ion. A protein wall blocks the straight passage of nucleic acids resulting in a bend in the path, which exposes the end of the DNA-RNA hybrid for addition of NTPs. The 5' end of the RNA catches a loop of protein, which prevents extension of the DNA-RNA hybrid beyond 9 bp, separating DNA from RNA. The exit path of the RNA passes beneath the rudder and beneath another loop of protein. Adapted from Cramer et al., [2000].

affinity that RNAP has for general regions of DNA [Gourse *et al.*, 2006] and also enables RNAP to recognise specific promoters. The discovery of sigma by Burgess and others in 1969, answered the question of how RNA polymerase is guided to the right promoters and prevented from binding unproductively to random sequences, which spurred decades of study into promoter recognition and transcription initiation. The holoenzyme binds duplex DNA and searches for a promoter site by forming transient hydrogen bonds with exposed hydrogen donor and acceptor groups of base pairs. RNAP slides along the DNA rather than repeatedly binding and dissociating from it. The –10 and –35 sequences upstream of transcription start sites can be found without the need for unwinding, therefore the search for the promoter region is rapid [Polyakov *et al.*, 1995].

Table 1.1. Subunits of bacterial RNA polymerase.

Subunit	***Size [aa]***	***Size kD***	***Gene***	***Function***
Alpha [α]	329	36.5	*rpoA*	Required for enzyme assembly. Interacts with some regulatory proteins. Involved in catalysis.
Beta [β]	1342	150.6	*rpoB*	Catalysis of RNA synthesis during initiation and elongation. Recognises terminator sequences. Binds ribonucleoside 5' –triphosphates.
Beta' [β']	1407	155.2	*rpoC*	Binds to the DNA template and the σ subunit.
Sigma [σ]*, e.g. σ^{70}	613	70.3	*rpoD**	Required for initiation of transcription. Recognises promoter sequences.
Omega [ω]	91	10.2	*rpoZ*	Restores denatured RNA polymerase *in vitro* to its fully functional form.

* *Alternative sigma factors and corresponding genes are listed in Table 1.2.*

The division of labour that exists between the core enzyme and that of the sigma factor, suggests that there is more than one type of sigma factor. In *E. coli*, there are several alternate factors, each specific for a different class of promoter [Table 1.2.]. The major class of bacterial RNAP holoenzymes contain a σ subunit, which belongs to the σ^{70} group and binds to the promoters of housekeeping genes.

The alternative major bacterial RNAP holoenzyme contains an unrelated sigma subunit, σ^{54} and the method by which transcription is initiated by σ^{70} and σ^{54} vary mechanistically [Wigneshweraraj *et al.*, 2005]. Genes under the control of σ^{54} are induced under specific conditions for example, during ammonium limited growth in *Klebsiella oxytoca*, phosphorylated NtrC protein activates transcription by binding to sequences upstream of the promoter binding sites. Promoters under NtrC protein control require σ^{54} [also known as σ^{N}], which is recruited to activate *nasFEDCBA* operon expression, which is subsequently subject to further regulation of gene expression by an antitermination mechanism [Chai *et al.*, 1998], [section 1.5.1.4].

There are five additional sigma factors in *E. coli* that recognise alternate sets of promoters: σ^{32}, σ^{E}, σ^{S}, σ^{F} and σ^{Fec}. These are activated in response to various environmental stresses, for example, σ^{32} promoters are induced in response to heat shock. A generally accepted model for the heat shock response is that due to the increased levels of σ^{32} in response to heat shock, sigma switches from the major σ^{70} subunit to heat shock responsive σ^{32}, directing RNAP to the heat shock genes [Wang and DeHaseth, 2003]. The *rpoE* gene, which encodes σ^{E} has been shown to be essential for bacterial growth at high temperatures [Hiratsu *et al.*, 1995]. Sigma^{E} is also induced under extra-cytoplasmic stress [reviewed in Gourse *et al.*, 2006]. Sigma^{S}, turns on stationary phase promoters and σ^{F} is used for flagellum-related functions. Sigma^{FecI} factor induces promoters involved in iron transport [Gourse *et al.*, 2006].

Although a common feature amongst all promoter regions is that they are located at the same distances from the transcription start point, each sigma factor causes RNAP to initiate from an exclusive set of promoters. Promoter sequences are found upstream of the region that codes for protein, and consist of short conserved DNA sequences [Section 1.4.1]. It is estimated that the complete chromosome of *E. coli* contains approximately 2000 promoters [McClure, 1985].

Table 1.2. The Sigma Factors Found in E.coli

Sigma Factor	***Gene***	***Function***
σ^{70}	*rpoD*	Housekeeping functions
σ^{54}	*rpoN*	Nitrogen-regulated
σ^{32}	*rpoH*	Heat shock gene transcription
σ^{E}	*rpoE*	Initiates an extra-cytoplasmic stress response
σ^{S}	*rpoS*	Gene expression in stationary phase cells
σ^{F}	*rpoF*	Expression of flagellar operons
σ^{FecI}	*fecI*	Regulates the *fec* genes for iron dicitrate transport

Sequence analysis of the promoters have shown that each set has a unique sequence element and therefore guarantees that it is recognised only by the RNAP driven by the appropriate sigma factor thus allowing mutually exclusive, promoter specific transcription initiation. However, a recent study showed that there is functional overlap between sigma factors, such as σ^{32} and σ^{70} [Wade *et al.*, 2006]. In some cases the overlap is extensive, in particular with σ^{E} and σ^{70}, suggesting that an important feature of how transcription is regulated in bacteria is the augmentation

of transcription from σ^{70}, thus allowing expression of genes under multiple growth conditions, and generating more complex regulatory patterns.

1.4 Bacterial Transcription

Bacterial transcription involves three stages, initiation, elongation and termination, and all of these processes are subject to varied intricate regulation enabling the expression of modified proteomes to meet the growth need of the bacterium.

1.4.1 Initiation

For the initiation of transcription, the σ subunit of RNAP holoenzyme must first recognise specific promoter sequences within the template DNA to allow the polymerase to initiate transcription. For example, σ^{70} promoters recognise two sequence motifs, -35 box [TTGACA] and -10 box [TATAAT]. Strong promoters have sequences that closely align to the consensus sequence, whereas weak promoter sequences have substituted bases within these regions. Initiation begins when RNAPσ^{70} binds DNA at the -35 box and forms a closed promoter complex. The double helix dissociates at -10 to form an open promoter complex. The transition from a closed promoter complex to an open promoter complex is a key event in transcription and marks the shift from the initiation stage to the elongation stage, which commences with the formation of the first phosphodiester bond in the nascent RNA. Finally, σ^{70} dissociates once a 7-12 nucleotide RNA molecule is synthesised.

1.4.2 Elongation

During elongation RNAP moves along the DNA and sequentially synthesises the RNA chain, by incorporating ribonucleotides to the 3' end of the chain. The elongation complex is stabilised by protein-protein interactions, and contact

between the template DNA and growing RNA transcript is made for 8-10 nucleotides. Elongation occurs at a rate of approximately 50 nucleotides per second, covering a distance of 170 Å.

The first structural study of *E. coli* core RNAP, carried out by Polyakov and others [1995] resolved the core polymerase to ~23 Å resolution by 3D reconstruction from electron micrographs of flattened helical crystals. This study revealed that the core RNAP undergoes a dramatic conformational change relative to the holoenzyme, such that the core enzyme resembled the previously determined structure of yeast RNAP II [Darst *et al.*, 1991]. Since both RNAP core and yeast RNAP II catalyse the processive elongation of RNA, where they remain bound to the DNA between each cycle of nucleotide addition, it was proposed that this structure was indicative of an 'elongation' conformation. The structure of the RNAP core, like all other known structures of RNA polymerases exhibits a thumb like projection, but whereas the holoenzyme thumb forms one wall of a ~25 Å wide groove, that of the core curls around to contact the opposite wall of the channel, forming a closed ring similar to that of yeast RNAP II [Polyakov *et al.*, 1995; Darst *et al.*, 1991] locking the RNAP core onto the DNA.

1.4.3 Termination of Transcription

Termination of transcription occurs once a terminator sequence is transcribed. Terminator sequences consist of a region of dyad symmetry, which can fold into stable stem loop structures and signal the RNAP to terminate transcription. There are two categories of terminators, these are either *Rho*-independent, [sometimes referred to as intrinsic termination], or *Rho*-dependent transcription termination.

1.4.3.1 Rho-independent Termination

Rho-independent terminator sequences contain two sequence motifs, which are necessary for termination of transcription; an RNA sequence that is able to form a

stable stem-loop structure, followed by a run of 9 uridine [U] residues. The sequence forming the stem-loop, which once transcribed, causes RNAP to stall. At this pause site, the paused stem-loop contacts a flexible domain on RNAP called the flap, which forms a critical part of a hairpin-interaction site on the enzyme [Wang *et al.*, 1997]. The second sequence motif consists of a run of residues immediately downstream of this. This is thought to promote dissociation of the RNAP by destabilising the attachment of the transcript to the template [Yarnell and Roberts 1999]. Although the precise mechanism of transcription termination is unclear, and there is no direct evidence of how termination is achieved at this point, it seems highly likely that when the terminator sequence is transcribed, a thermodynamically stable RNA-RNA base paired stem-loop structure is formed and favoured over the DNA-RNA base pairing that would otherwise occur [Komissarova *et al.*, 2002], diminishing the number of contacts made between the template DNA and RNA transcript, which weakens the overall stability of the complex. In addition, as the remaining contacts between the DNA: RNA hybrid occurs between U residues in the RNA and A residues in the DNA, dissociation of the complex is favoured due to the weak base pairing between A and U. The RNA molecule dissociates from the DNA, still within the transcription bubble, the DNA template subsequently rejoins to its partner to reform the duplex, and the core enzyme, having much less affinity for double stranded DNA releases the DNA.

1.4.3.2 The Role of Rho

A transcribed RNA sequence complete with a stable stem loop structure followed by a run of U residues is usually sufficient for the termination event to occur without the requirement for additional factors. In some instances however, such as in the absence of U residues, termination can only be achieved by the participation of a terminator factor. One such factor, *Rho* [ρ], binds to the RNA transcript at a *Rho* utilisation site [*rut*] and loads the RNA into its central channel before adopting a conformation able to translocate along the nascent RNA, utilising ATP and seeking out non-contiguous structural features within the transcription bubble

which signal termination. *Rho* is a ring shaped hexameric helicase that is an essential protein in *E. coli*, although there are few *Rho*-dependent terminators in the *E. coli* genome. Each monomer is 46kD and *Rho* is active as a hexameric protein of 275 kDa [Figure 1.2.]. Its six subunits are identical, which form a six-fold asymmetric ring, each having an ATP binding site located at the interface of the C terminal domains [Adelman *et al.*, 2006]. Each ATP binding site is composed of WalkerA/WalkerB motifs which are reported to be found in RecA – like proteins, which utilise NTP to perform mechanical work in the cell. Its mechanism of action is thought to be that of an ancillary factor to RNA polymerase which breaks the RNA-DNA hybrid by utilising ATP to dislodge the nascent RNA from the holoenzyme within the transcription bubble [Adelman *et al.*, 2006].

Although the initiation of transcription is the main target for regulation, numerous strategies have evolved that allows regulation of almost every step involved in the expression of non-constitutive genes. One such mechanism is to regulate the termination event of transcription by an antitermination mechanism. It is this strategy of transcriptional regulation in *P. aeruginosa*, which is the focus of this thesis and will be discussed further.

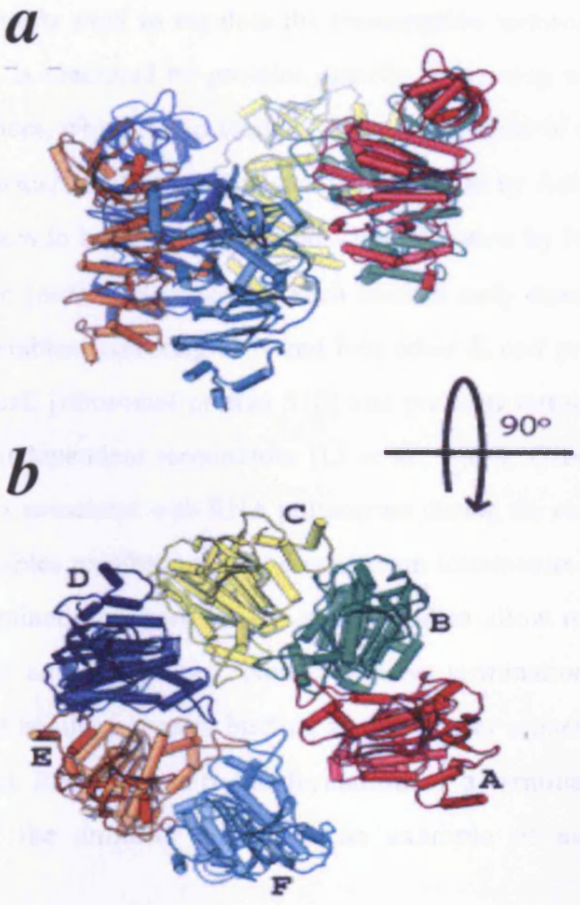


Figure 1.2. The Hexameric Structure of Rho

[a] The six identical subunits pack into an open ring. [b] Top down view. The alternating subunits are labelled A -F. [Figure adapted from Skordalakes and Berger, 2003].

1.5 Antitermination

A mechanism commonly used to regulate the transcription termination event is antitermination, that is mediated by proteins directly interacting with RNAP or leader mRNA sequences, which lie upstream of the coding region of the gene. The paradigm system in which antitermination was first described by Adhya and others [1974], is the N protein in bacteriophage λ and antitermination by N is dependent upon an N utilisation [*nut*] sites present in each lambda early operon. At these sites, a complex assembles consisting of N and four other *E. coli* proteins, NusA, NusB, NusG and NusE [ribosomal protein S10] and prevents termination at rho-dependent and rho-independent terminators [Li *et al.*, 1992; Greenblatt *et al.*, 1998]. This complex associates with RNA polymerase during the elongation stage and consequently enables readthrough past downstream terminators. This type of transcription antitermination, where RNAP is modified to allow readthrough, is known as processive antitermination. Non-processive termination, describes a mechanism where an ancillary protein binding at a particular sequence within the transcribed 5' leader RNA, prevents the formation of a terminator structure. Antitermination of the amidase operon is an example of non-processive antitermination.

1.5.1 Classic Systems Controlled by Non-processive Antitermination

1.5.1.1 The bgl Operon of E. coli

The *bgl* operon is one of four β-glucoside utilisation systems found in *E. coli* and the genes *bglG*, *bglF*, and *bglB* are involved in the uptake and utilisation of the aromatic beta-glucosides, salicin and arbutin [Boss *et al.*, 1999]. The antiterminator protein BglG controls transcription termination of the *bgl* operon [Figure 1.3.], [Amster-Choder and Wright, 1993].

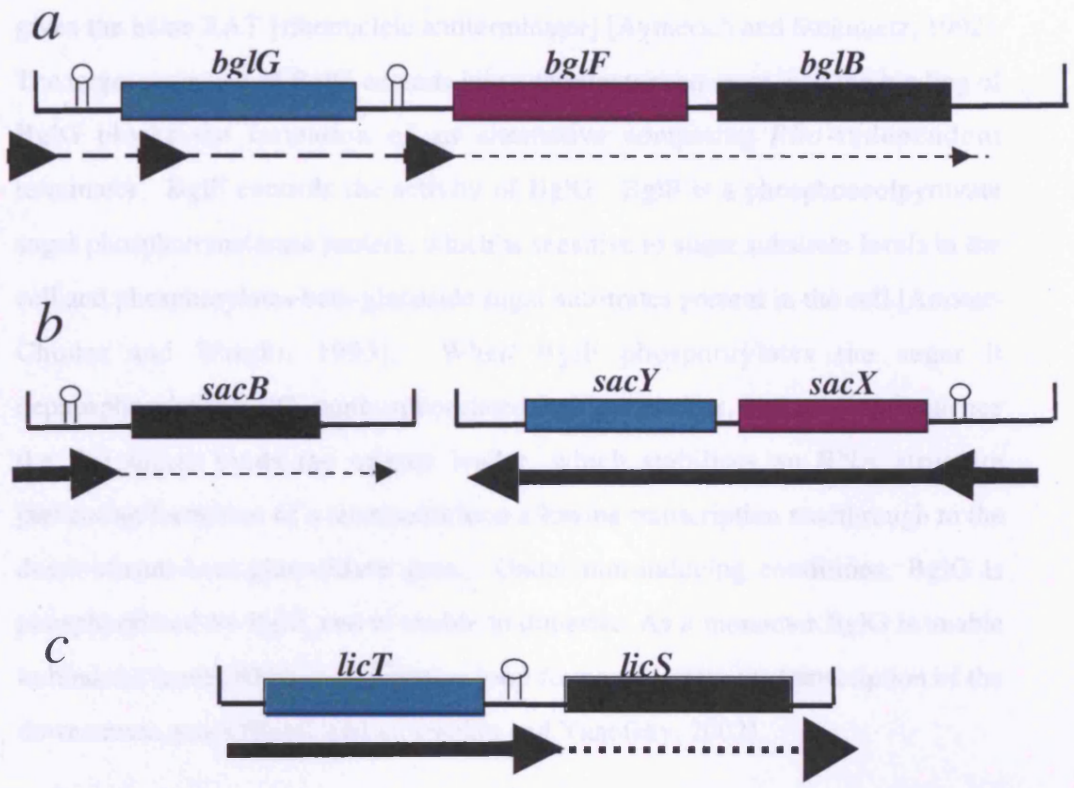


Figure 1.3. Organisation of the Genes of the Bgl, Sac and Lic Operons. *Antitermination of all three operons results in transcriptional readthrough and production of full length transcripts when the antiterminator protein [Teal] binds to a RAT sequence upstream of an intrinsic terminator.[a] BglG binds to a 32 nt RAT sequence immediately upstream of the second stem loop preventing the formation of a terminator loop. [b] In the presence of sucrose, activated SacY protein binds to a RAT sequence at the 5' end of an intrinsic terminator allowing transcription of sacB. [c] LicT protein binds to a sequence upstream of the terminator and thus allows production licS transcripts.*

Teal and purple denote the regulatory genes of the operons: Teal denotes the antiterminator gene of the operon and purple the negative regulator. Black denotes the structural genes of the operon.

BglG is able to bind to a 32 nucleotide target sequence within the leader mRNA immediately upstream of the first terminator sequence. This sequence has been given the name RAT [ribonucleic antiterminator] [Aymerich and Steinmetz, 1992]. The target sequence of BglG extends into a terminator sequence, and the binding of BglG blocks the formation of an alternative competing *Rho*-independent terminator. BglF controls the activity of BglG. BglF is a phosphoenolpyruvate sugar phosphotransferase protein, which is sensitive to sugar substrate levels in the cell and phosphorylates beta-glucoside sugar substrates present in the cell [Amster-Choder and Wright, 1993]. When BglF phosphorylates the sugar it dephosphorylates BglG, unphosphorylated BglG dimerises, and as a consequence the Bgl dimer binds the operon leader, which stabilises an RNA structure preventing formation of a terminator loop allowing transcription readthrough to the down-stream beta-glucosidase gene. Under non-inducing conditions, BglG is phosphorylated by BglF, and is unable to dimerise. As a monomer BglG is unable to bind the leader RNA, a termination loop forms and prevents transcription of the downstream genes [Reviewed in; Henkin and Yanofsky, 2002].

1.5.1.2 The sac Operon of Bacillus subtilis

The *B. subtilis* antiterminator protein SacY, mediates the induction of the SacB levansucrase in the presence of the inducer sucrose [Figure 1.3. (b)]. Transcription of *sacB* initiates constitutively at the promoter, however, in the absence of sucrose, transcription is prematurely terminated at a *Rho*-independent terminator, located upstream of the *sacB* gene [Declerck *et al.*, 2002]. In the presence of sucrose, SacY is activated and subsequently binds to the nascent RNA at the canonical RAT sequence. This RAT sequence partially overlaps the 5' end of a *Rho*-independent terminator. This *Rho*-independent terminator is intrinsically more stable than the antiterminator stem-loop, and thus is more likely to form in the absence of an inducer. Activated SacY binds to and stabilises the RAT structure, thereby preventing the formation of the terminator thus transcription proceeds to the

downstream gene. SacY is a 278-residue protein and the antitermination activity involves the first 55 residues [co-antiterminator domain; CAT] and antitermination activity by CAT is modulated by a phosphorylation/dephosphorylation mechanism involving the remaining 223 residues of the protein [Declerck *et al.*, 2002] [Figure 1.4.]. Two homologous regulatory domains regulate activity of CAT; PRDI and PRDII [phosphotransferase regulation domains], which are reversibly phosphorylated on conserved histidine residues [Manival *et al.*, 1997; Stulke *et al.*, 1998]. Phosphorylation of the PRDs affects the overall structure of the protein, and consequently determines the ability of CAT binding to the RAT sequence.

1.5.1.3 The lic Operon of B. subtilis

The CAT domain is also found in the *B. subtilis* antiterminator protein LicT, which regulates expression of operons involved in β-glucoside metabolism [Figure 1.3.(c)], [Declerck et al., 1999]. LicT is made up of an N-terminal CAT domain, and the PRDI and PRDII systems. In an activated state, both PRDs form dimeric units, with the phosphorylation sites buried at the interface [Graille *et al.*, 2005]. The NMR solution structure of LicT/RNA complex [Yang *et al.*, 2002], shows that the LicT-CAT domain folds as a symmetric dimer [Figure 1.5.]. Preliminary NMR footprinting experiments suggested that the protein surface, which interacts with the RNA, is located on one side of the dimer and both of the monomers are involved in RNA recognition. LicT-CAT interacts with the minor groove of the double stranded region of the RNA containing a stem flanked by two asymmetric internal loops. Comparisons between the active and inactive crystal structures show that there are massive tertiary and quaternary rearrangements of the entire regulatory domain [Graille *et al.*, 2005]. In the inactive state, a wide swing movement of PRDII results in dimer opening bringing the hidden phosphorylation sites to the surface. This opening is accompanied by further structural rearrangements of PRDI interface.

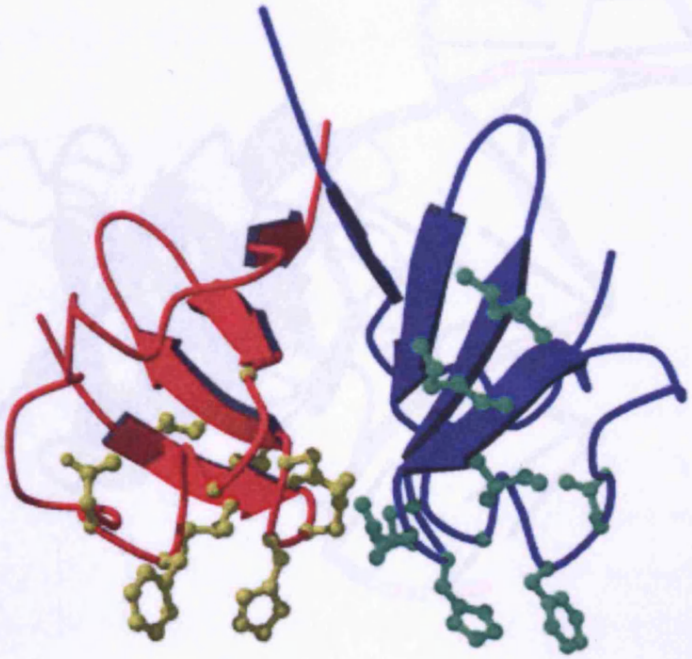


Figure 1.4. The Structure of the SacY Dimer.

The SacY [1-55] dimer interacts with a 29 nucleotide RAT sequence. SacY monomers represented in blue and red. Residues affected by RNA binding are shown in ball and stick representation in yellow and green on each monomer. Image adapted from Manival et al., [1997].

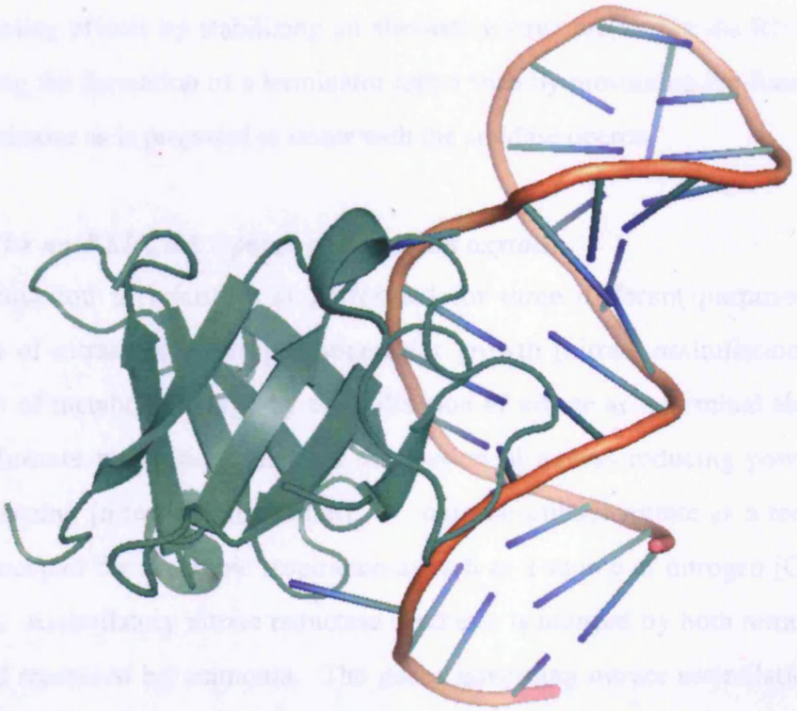


Figure. 1.5. Structure of the LicT-RNA Complex.

Activated, LicT bound to a 29 base RNA hairpin overlapping a terminator located in the 5` mRNA leader region of the target gene. The solution structure of the LicT RNA-binding domain, CAT, in complex with its RAT target by NMR spectroscopy. CAT is a beta-stranded homodimer that undergoes no important conformational changes upon complex formation. The interaction is mostly hydrophobic and stacking interactions, with the distorted minor groove of the hairpin stem that is interrupted by two asymmetric internal loops [Yang et al., 2002]. Image created from PDB file 1LIC using Pymol.

SacY, BglG and LicT belong to a 56 member family of structural homologues which have high sequence similarity [Aymerich and Steinmetz, 1992] [Schnetz *et al.*, 1996], [Manival *et al.*, 1997]. Similarities can be found between regulation of the amidase operon and *bgl* and *sac* systems up to the point that transcription is activated by an antitermination mechanism. However, both *bgl* and *sac* exert their antiterminating effects by stabilizing an alternative structure within the RNA and not allowing the formation of a terminator rather than by preventing the formation of the terminator as is proposed to occur with the amidase operon.

1.5.1.4 The nasFEDCBA Operon of Klebsiella oxytoca

Nitrate reduction in microbes is performed for three different purposes: the utilization of nitrate as a nitrogen source for growth [nitrate assimilation], the generation of metabolic energy by the utilisation of nitrate as a terminal electron acceptor [nitrate respiration], and the dissipation of excess reducing power for redox balancing [nitrate dissimilation]. *K. oxytoca* utilises nitrate as a terminal electron acceptor for anaerobic respiration as well as a source of nitrogen [Chai *et al.*, 1998]. Assimilatory nitrate reductase synthesis is induced by both nitrate and nitrite and repressed by ammonia. The genes governing nitrate assimilation are encoded in the *nasFEDCBA* operon of *K. oxytoca* [Figure 1.6.], [Goldman *et al.*, 1994]. During ammonium-limited growth, phosphorylated NtrC protein activates transcription from upstream binding sites. In addition to the global regulation of gene expression exerted by NtrC, the *nasFEDCBA* operon is subject to control via a transcription antitermination mechanism [Lin and Stewart, 1996]. The *nas* operon genes encode a nitrate and nitrite transport system. Nitrate is converted to nitrite then to ammonium by assimilatory nitrite reductase encoded by the *nasC* and *nasA* gene products and assimilatory nitrite reductase encoded by *nasB*. The *nasR* gene is the positive regulator of transcription of the genes *nasFEDCBA* and in the presence of either nitrate or nitrite is able to bind to a

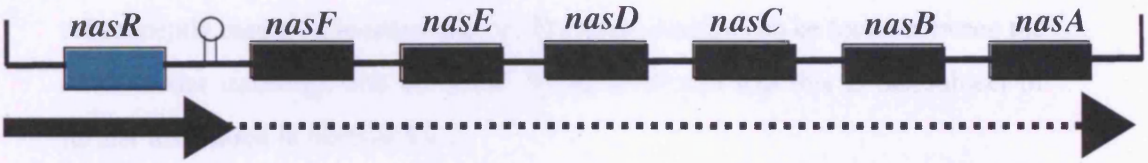


Figure 1.6. Organisation of the Nas operon.

The NasR gene encodes a transcription antiterminator protein which under inducing conditions, binds to a terminator sequence located in the nasF leader region and prevents transcription termination. The nasFED genes encode a nitrate and nitrite transport system and nasCBA encode nitrate and nitrite reductases.

terminator sequence located in the *nasF* upstream leader region to prevent transcription termination. Chai and Stewart [1999], showed that the *nasF* leader transcript has two stem-loop secondary structures – 1:2 and 3:4 – the *rho*-independent termination. Deletions into the 3:4 stem-loop results in constitutive *nasF* operon expression and deletions into the 1:2 stem-loop, which is essential in nitrate and nitrite induction, results in a non-inducible phenotype. According to Chai and Stewart [1998], it is likely that the 1:2 stem-loop serves as a scaffold for the nitrate-activated NasR protein, guiding it to the 3:4 stem-loop, which subsequently results in antitermination. Many similarities can be found between the *nasF* leader transcript and the *amiE* leader transcript and this is the subject of further discussion in Section 5.1.1.

All of the systems discussed here so far except for the *nasFEDCBA* operon are catabolic operons. Although the anabolic *nas* operon is involved in nitrate assimilation, it is included here due to its similar antitermination mechanism.

1.6 Amidase operon of Pseudomonas aeruginosa.

In 1959, a strain of *P. aeruginosa* [PAC1] was found to utilise acetamide and propionamide as sole sources of carbon and nitrogen by producing an inducible aliphatic amidase [Kelly and Clarke, 1962]. These amides were able to induce a chromosomally located aliphatic amidase, which upon induction, rapidly hydrolysed the amides to their corresponding organic acids and ammonia.

1.6.1 Substrates and Inducers of Amidase Expression

Early studies showed that the substrate/inducer profiles of the amidase system were distinct such that acetamide, propionamide, lactamide, and N-acetylacetamide were the best enzyme inducers and formamide, butyramide, and phenylacetamide the most active amide analogue co-repressors of the system [Table 1.3.], [Kelly and Clarke, 1962]. Formamide is a poor substrate, and although it can be hydrolysed by amidase constitutive mutants, it provides only a nitrogen source since one-carbon compounds cannot be assimilated by this species [Farin and Clarke, 1978]. The four-carbon amide, butyramide was found to repress amidase synthesis, competing with non-substrate inducers such as N-acetylacetamide, but it can be hydrolysed poorly by the amidase system [Farin and Clarke, 1978]. Lactamide, although not an amidase substrate, is able to induce enzyme synthesis, and thus serves as a gratuitous inducer. Amidase was shown to have hydrolase activity, transferase activity in the presence of hydroxylamine and some esterase activity. It was also shown that enzyme synthesis was subject to carbon catabolite repression and that carbon sources that supported highest growth rates caused the greatest repression [Brammar and Clarke, 1964].

Table 1.3. Amide-substrate inducer profile of the amidase operon

Amide	*Substrate*	*Inducer*	*Co-repressor Activity*
Formamide [$HCONH_2$]	+	-	+
Acetamide [CH_3CONH_2]	+ +	+ +	_
Propionamide [$CH_3CH_2CONH_2$	+ + +	+ + +	_
Butyramide [$CH_3CH_2CH_2CONH_2$]	+	_	+ + +
N-Acetylacetamide [$CH_3CONHCOCH_3$]	_	+ + +	
Lactamide [$CH_3CHOHCONH_2$]	_	+ + +	_

1.6.2 Isolation of the Genes of the Amidase Operon

The amidase genes were isolated from the constitutive mutant PAC433 from a bacteriophage lambda vector 722 gene library [Drew *et al.*, 1980]. A 5.3 kb *Hind*III/*Sal*I restriction fragment containing the amidase operon genes was sub-cloned into pBR322 to generate plasmid pJB950 [Clarke *et al.*, 1981]. This procedure was also used to sub-clone the wild type amidase operon genes from PAC1 to generate plasmid pAS20 [Wilson and Drew, 1991]. The complete fragment containing the amidase genes has been sequenced and five genes identified, *amiEBCRS*, which appear to form the entire amidase operon [Section 1.7.6].

1.7 Genes of the Amidase Operon

1.7.1 AmiE

The *amiE* gene encodes the amidase enzyme. Amidase is a hexameric protein, consisting of six identical subunits with a molecular mass of 38.4 kDa. *In vivo*, amidase catalyses the hydrolysis of amides, *in vitro* however, amidase is able to catalyse a number of different reactions [Table 1.4]. The amide transferase reaction can be used to measure the amount of amidase activity in intact cells using acetamide as a substrate [Brammer and Clarke 1964]. Recent sequence analysis has shown that the amidase belongs to the carbon-nitrogen hydrolase family and this family includes nitrilases and biotinases. Amidases have been detected in *Geobacillus stearothermophilus*, *Rhodococcus erythropolis*, *Bradyrhizobium japonicum* and *Helicobacter pylori* and have a greater than 90% sequence identity [Drew and Haq, 2002].

Table 1.4. Reactions Catalysed by Amidase

Enzyme Activity	***Chemical Reaction***
Amide hydrolysis	$RCONH_2 + H_2O \Rightarrow RCO_2^- + NH_4^+$
Ester hydrolysis	$RCO_2R' + H_20 \Rightarrow RCO_2^- + H^+ + ROH$
Amide Transferase	$RCONH_2 + NH_2OH \Rightarrow RCONHOH + NH_4^+$
Ester transferase	$RCO_2R + NH_2OH \Rightarrow RCONHOH + ROH$
Acid transferase	$RCO_2^- + NH_2OH + H^+ \Rightarrow RCONHOH + H_2O$

1.7.2 AmiB and AmiS

Complete sequencing of the pAS20 insert revealed two new ORF's downstream of *amiE*. These were identified as *amiB* and *amiS*. The *amiB* gene encodes a 42 kDa protein and studies have shown that deletion of *amiB* has no effect on the inducibility of amidase operon. Hydropathy analysis of the AmiB amino acid sequence indicated that this was a substantially hydrophilic protein and therefore its cellular location highly likely to be cytoplasmic [Wilson and Drew 1991]. The *amiS* gene encodes an 18 kDa protein, with a high content of hydrophobic residues. Hydropathy analysis suggests the presence of six transmembrane helices and is thus regarded as a putative integral membrane protein. Both AmiB and AmiS bear a resemblance to components of an ABC transporter system [Wilson *et al.*, 1995], although no role has been assigned to AmiB and AmiS thus far.

1.7.3 AmiC

The *amiC* gene was identified as the negative regulator of the system when *in vitro* constructed deletions and insertions upstream of *amiR* led to constitutive amidase expression. This was further supported by the findings that a broad host range expression vector carrying *amiC* repressed amidase expression in a set of amidase-constitutive mutants [Wilson and Drew, 1991]. Over expression of *amiC* in the wild-type strain PAC1 results in a non-inducible phenotype [Wilson and Drew, 1991]. The crystal structure of AmiC protein has been determined to a resolution of 2.1Å, which showed AmiC to be a two-domain protein [Figure 1.7]. Each domain consists of β-α-β topology linked by a hinge region [Pearl *et al.*, 1994]. The overall fold of AmiC is related to the periplasmic small molecule binding proteins, such as LivJ [Oxender *et al.*, 1980], the branch chain amino acid leucine, isoleucine and valine binding protein [Pearl *et al.*, 1994]. The crystal structure of the AmiC-acetamide complex defined the amide-binding site, which occurs at the

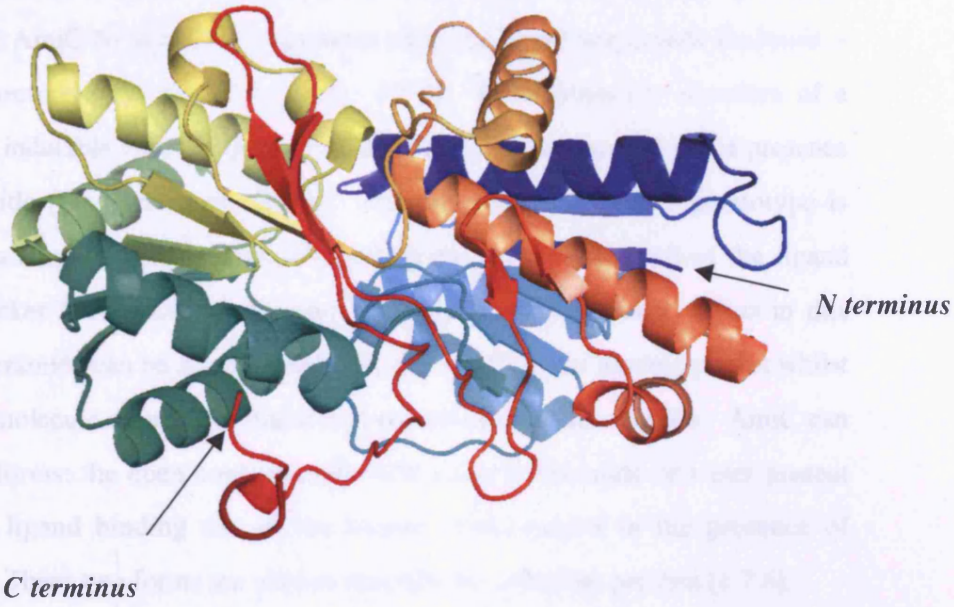


Figure 1.7. Structure of AmiC.

The 2.1Å crystal structure showing the two-domains of the AmiC protein. Each domain consists of β-α-β topology linked by a hinge region. Image created from PDB file 1QNL using the 3D structure viewer Pymol.

interface of the two domains. Binding of acetamide at this site stabilises the two domains into a 'closed' conformation. AmiC is also able to bind non-inducing amides such as butyramide. Butyramide, which has a longer 3-carbon chain, is not easily accommodated by AmiC at the binding site and binds with approximately 100-fold lower affinity than acetamide [Wilson *et al.*, 1993]. The structure of AmiC-butyramide as seen in the AmiC/AmiR-butyramide complex [Figure 1.8], is similar to that of AmiC-acetamide except for a slight inter-domain movement between the AmiC N- and CTD to accommodate the larger butyramide molecule – the open conformation [O' Hara *et al.*, 1999]. In addition the structure of a butyramide inducible AmiC mutant [PAC181] has been determined in the presence of butyramide [O' Hara *et al.*, 2000]. The butyramide inducible phenotype is caused by a single T106N mutation in which part of the side wall of the ligand binding pocket is removed to increase the volume of the pocket. Thus in this mutant butyramide can be accommodated in the AmiC ligand binding pocket whilst the AmiC molecule adopts the [inducing] closed-down conformation. AmiC can adopt two forms: the open conformation with either butyramide or water present within the ligand binding site or the closed conformation in the presence of acetamide. These two forms are used to describe the induction process [1.7.6].

1.7.4 AmiR

The *amiR* gene was initially shown to encode a positive regulator of amidase operon expression by the isolation of temperature sensitive mutants, which were amidase negative at high temperature [Farin and Clarke, 1978]. The location of *amiR* was determined initially by co-transduction [Brammar *et al.*, 1967], and subsequently by deletion mapping and sub-cloning restriction enzyme fragments [Cousens *et al.*, 1987]. The mechanism of regulation identified as AmiR-dependent transcription antitermination identified by the use of a terminator deletion [Drew and Lowe, 1989]. Attempts to isolate and investigate AmiR were unsuccessful

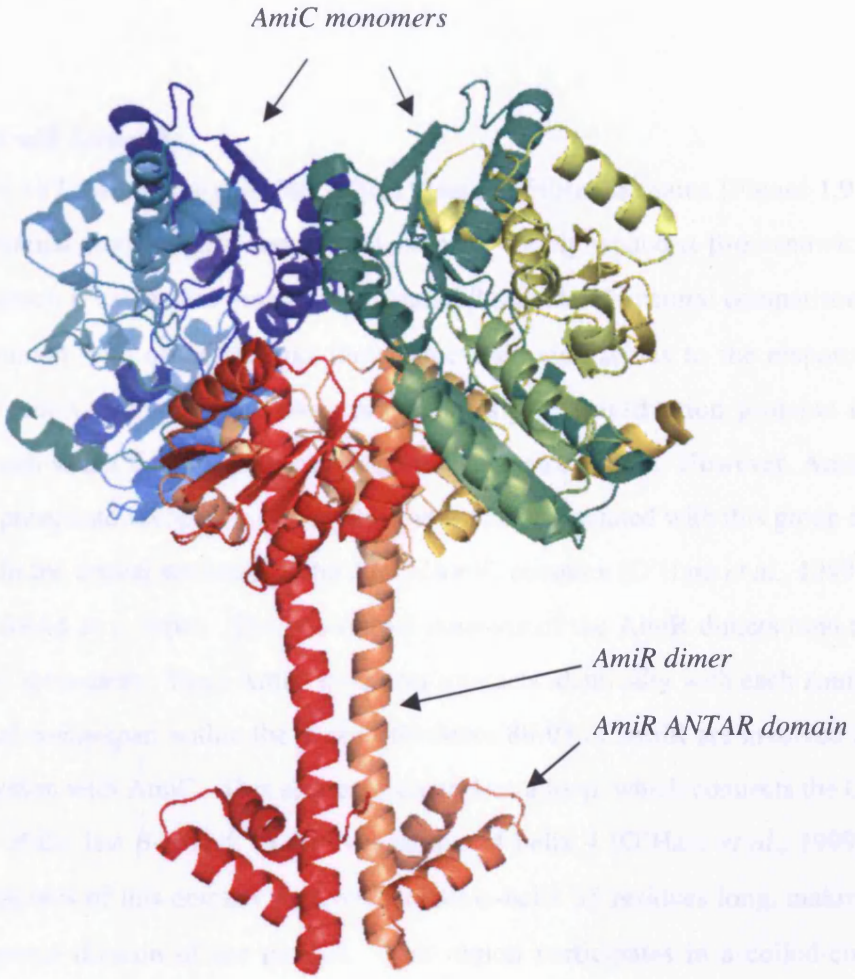


Figure 1.8. The AmiC/AmiR Complex

The N terminal domains of the AmiR dimers bind to two AmiC monomers. The two AmiC monomers within the complex can be seen in green and blue [chain a and b respectively] and the AmiR dimer in red and orange [chain d and e respectively]. Image created from PDB file 1QO0, using Pymol.

until the AmiC-AmiR-butyramide complex was identified as a stable fraction for isolation purposes. Structural determination of the complex followed [O'Hara *et al.*, 1999], along with the proposed mechanism of operon regulation [Norman *et al.*, 2000].

1.7.4.1 AmiR Structure

AmiR is a 197-residue polypeptide, which consist of three domains [Figure 1.9]. The N-terminal domain, [residues 1-128], forms a doubly wound α-β-α sandwich domain, which is assembled around five parallel β-strands. Structural comparisons of this domain with other proteins have uncovered similarities to the response regulator receiver domains of two component signal transduction proteins in bacteria such as the *E.coli* chemotaxis response regulator, CheY. However, AmiR lacks the phosphate acceptor residue, which is usually associated with this group of proteins. In the crystal structure of the AmiR/AmiC complex [O'Hara *et al.*, 1999], AmiR is found as a dimer. The N terminal domains of the AmiR dimers bind to two AmiC monomers. Each AmiC monomer interacts identically with each AmiR N-terminal counterpart within the dimer. Residues 88-93 of AmiR are involved in the interaction with AmiC. This sequence comprises a loop, which connects the C-terminus of the last β-strand, to the N-terminus of helix 4 [O'Hara *et al.*, 1999]. The C-terminus of this domain stretches into an α-helix 55 residues long, making up the central domain of the protein. This region participates in a coiled-coil interaction with the equivalent helix of the second AmiR monomer, and the interface between the two is principally a hydrophobic attraction. The CTD of AmiR is thought to contain the antitermination activity, since deletion of either the last helix or the last two helices results in loss of antitermination activity *in vivo* [O' Hara *et al.*, 1999]. Homology searches and structural predictions have identified the AmiR 3-helix bundle as the prototype for a > 40 member RNA-binding ANTAR family [Shu and Zhulin, 2002]. Family members are spread through the bacterial Kingdom and include proteins from important pathogens and symbionts. In each

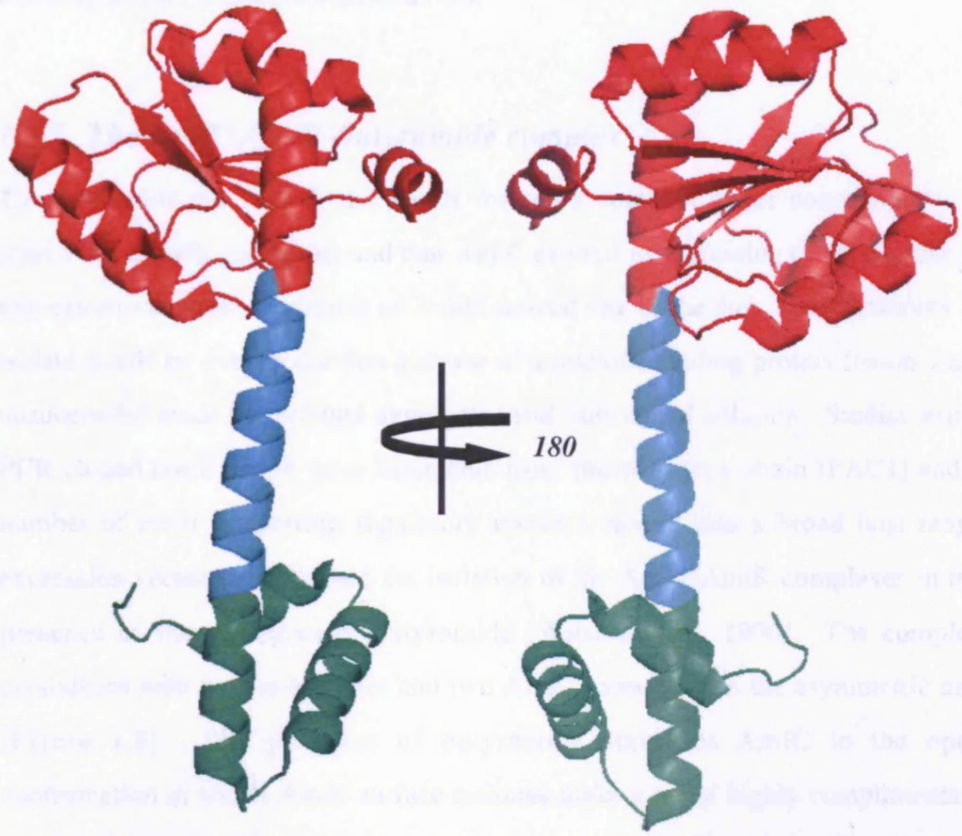


Figure 1.9. Domain architecture of AmiR.

Residues 1-128 in red showing the globular alpha-ß-alpha sandwich domain. Residues129-147 in cyan: a long alpha helix, involved in coiled coil interaction with the equivalent helix of the AmiR dimer. Residues 148- 195 [Green]; The C-terminal ANTAR domain, which forms a 3-helix bundle and a putative RNA binding domain. Image created from PDB file 1QO0; chain D, using Pymol.

case it is presumed that the identified ORF has the ANTAR domain as an RNA-binding response regulator output domain and that regulation occurs by a type of transcription antitermination mechanism.

1.7.5 The AmiC/AmiR-butyramide complex

The realisation that AmiC and AmiR formed a complex under non-inducing or repressing growth conditions and that AmiC exerted its repressing effect on operon expression via steric hindrance of AmiR arrived late in the day. Early attempts to isolate AmiR by overproduction and use of a maltose-binding protein fusion were unsuccessful since the product aggregates and falls out of solution. Studies using PCR cloned *amiC*, *amiR* gene fragments from the wild type strain [PAC1] and a number of other interesting regulatory mutants cloned into a broad host range expression vector first allowed the isolation of the AmiC/AmiR complexes in the presence of the co-repressor butyramide [Wilson *et al.*, 1996]. The complex crystallises with an AmiR dimer and two AmiC monomers in the asymmetric unit [Figure 1.8]. The presence of butyramide stabilises AmiC in the open conformation in which AmiC surface residues make a set of highly complimentary intereactions with the top head of the AmiR dimer. The formation of these interactions are absolutely dependent on the AmiC open conformation [O' Hara *et al.*, 1999]. Each AmiC monomer interacts identically with each AmiR N-terminal counterpart within the dimer.

1.7.6 Regulation of the Amidase Operon is via a Version of a Two Component Signal Transduction System involving AmiC and AmiR

Gene expression of the amidase operon is regulated via a transcription antitermination mechanism orchestrated by AmiC and AmiR [Drew and Lowe, 1989: Wilson and Drew, 1991]. AmiR is the positive regulator of the system. Under inducing conditions, free AmiR is able to bind to the RNA sequence upstream of the *Rho*-independent terminator, and prevent the terminator stem-loop from forming. The precise mechanism of how AmiR executes its antitermination activity is currently not understood. Under non-inducing conditions, the activity of AmiR is sterically hindered by the negative regulator AmiC [Norman *et al.*, 2000]. Under these conditions, AmiR is unable to antiterminate transcription, and transcription partially terminates when the terminator sequence is transcribed. AmiC controls the activity of AmiR by complex formation and thus prevents it from binding RNA. AmiC acts as the ligand sensor of the pair, and binds inducing amides. Thus in the presence of butyramide or absence of inducing amides, AmiC adopts the 'open' conformation and the AmiC/AmiR complex forms. In the presence of acetamide, AmiC adopts the 'closed' conformation, which breaks the highly complimentary interaction between the two AmiC monomers and the 'top' surface of the AmiR dimer and the complex dissociates. In this state, AmiR is released by AmiC, and is free to bind to the leader region sequences to allow RNA polymerase to read through the terminator and proceed to transcribing the downstream genes. Thus, AmiC and AmiR are the regulatory proteins of the amidase operon which form a two component signal transduction system in *P. aeruginosa.*

The majority of bacterial transcription response regulators contain two characteristic structural motifs consisting usually of an N-terminal receiver domain linked to a C-terminal output domain [reviewed in Parkinson, 1993]. The

mechanism exhibited by the amidase systems differs from most of two-component signal transduction systems found in bacteria, whereby covalent modifcation by phosphorylation is used to transmit the signal from one component to the next [reviewed in Parkinson, 1993]. In the amidase system, AmiR remains unmodified, but is released from AmiC after a conformational change induced by ligand binding.

1.7.7 Operon Leader Region

Expression of the amidase operon genes is constitutively initiated at a strong *E. coli* like σ^{70} promoter, located 150bp before the *amiE* gene translation initiation codon. The promoter is followed by a short open reading frame [ORF], a *Rho*-independent terminator sequence, a conventional ribosome binding site [Shine-Dalgarno sequence] and the *amiE* gene [Figures 1.10 and 1.11]. *In vivo* titration studies have previously shown that antitermination occurs by a direct interaction of AmiR with the leader RNA transcript [Wilson *et al.*, 1996; Norman *et al.*, 2000]. Under non-inducing conditions transcription of the upstream region produces a leader mRNA of no apparent function, which contains an ORF of 35 amino acids [Figure 1.11].

RNA folding studies of the leader sequence show the formation of a stable structure with an overall free energy of formation of –43.7 kcal/mol. The sequence contains a short leader, a large RNA hairpin [residues 5-37] with a free energy of formation of –16.1 kcal/mol, a central unstructured region, and finally the terminator hairpin [residues 66-95], with a free energy of formation of –23.3 kcal/mol, followed by a run of U residues [Figure 1.12]. Previous studies showed that a shortened leader [36-75] was able to titrate out AmiR antitermination activity [Wilson and Drew, 1996] implying that sequences in the first hairpin were not necessary for antitermination reaction. The central unstructured region was investigated [residues 66-95] by the construction of mutations and deletions. Analysis of these

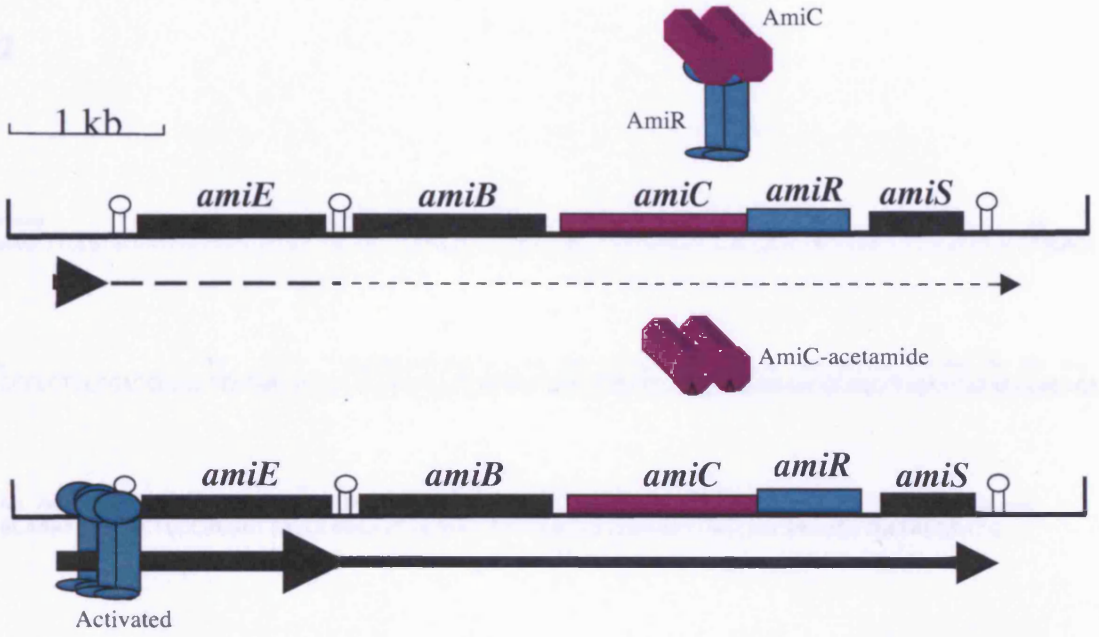
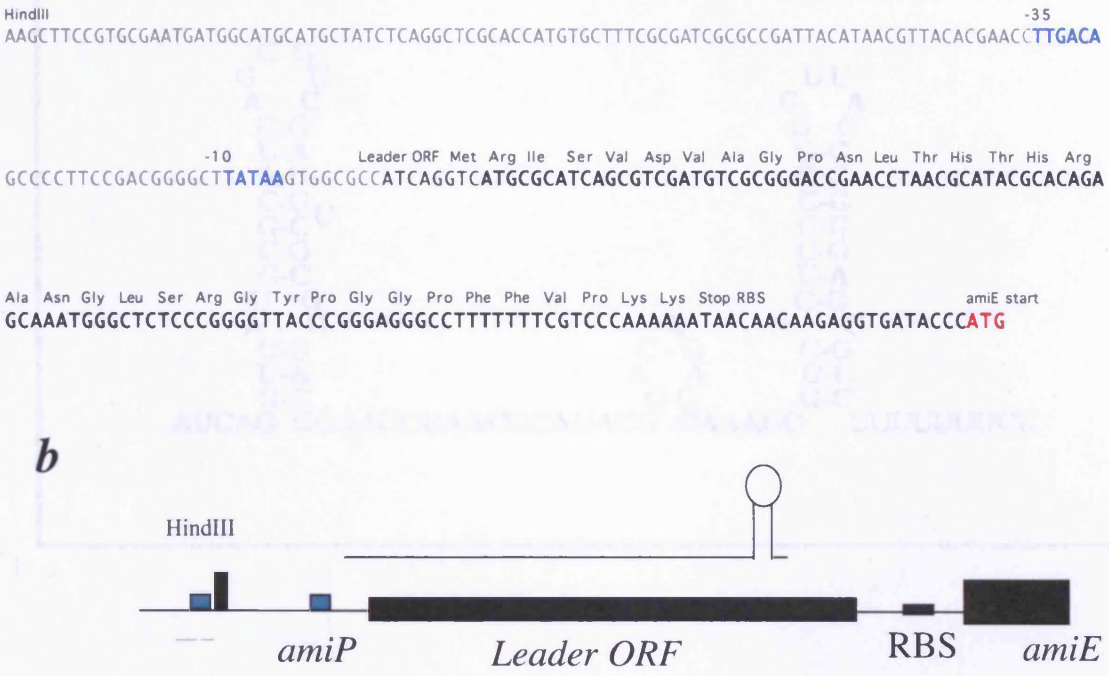


Figure 1.10. Transcription regulation of the amidase operon.

[a] Under non-inducing conditions, basal levels of amiE transcripts are produced and smaller amounts of amiB, amiC, amiR and amiS are produced. In this state, two AmiC monomers associate with an AmiR dimer and activity of AmiR is prevented. The majority of transcripts terminate at an intrinsic terminator [denoted by the first stem loop structure]. [b] Under inducing conditions, AmiC releases AmiR. AmiR is then able to bind the 5' leader RNA and prevents termination, thus allowing read through into the remaining genes of the operon.

a



b

Figure 1.11. DNA Sequence and features of the amidase operon leader region. *[a] The DNA sequence between the promoter and the amiE translation start point contains a short ORF of unknown function which includes an E. coli like rho-independent transcription terminator. [b] Features of the operon leader region showing the promoter, leader ORF, amiE RBS and gene locations.*

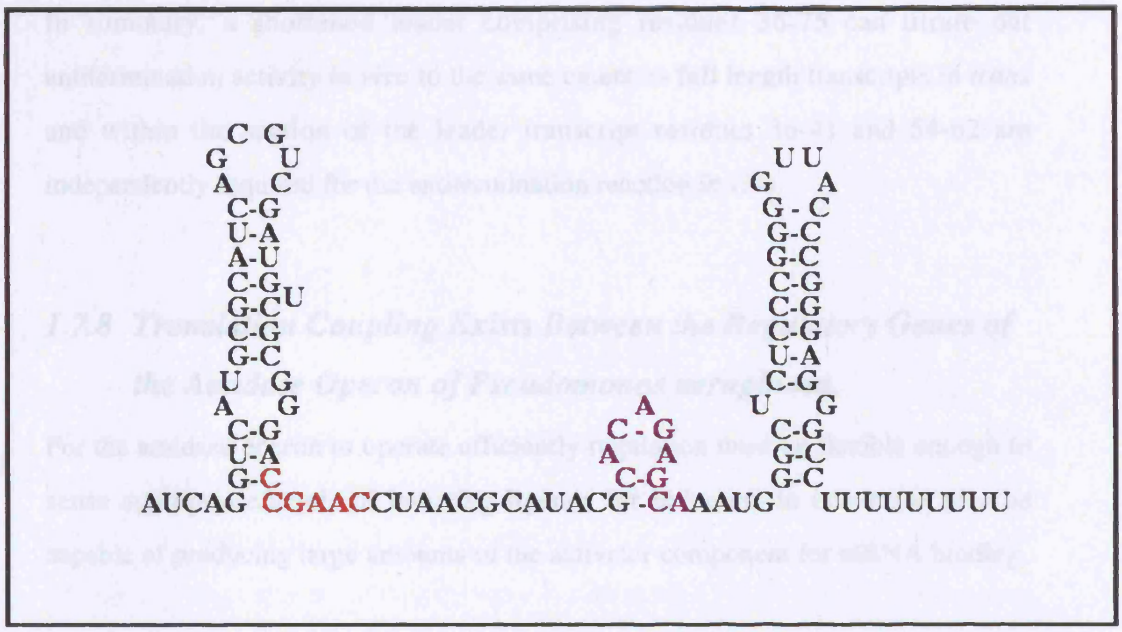


Figure 1.12. RNA fold of the leader region of the amidase operon.

The L [Left] and R [Right] recognition sequences are coloured red and purple respectively. The leader sequence illustration shows that it can form into a stable structure with an overall free energy of formation of –43.7 kcal/mol. RNA fold generated using an RNA folding programme; Mfold [Zuker, 2003].

showed two parts of the sequence CCGAAC [36-41] - referred to as the 'L' [left] recognition region and CACAGAGCA [54-62] - referred to the 'R' [right] recognition region. These were found to be critical to the antitermination reaction, changes to which abolished antitermination activity *in vivo*.

In summary, a shortened leader comprising residues 36-75 can titrate out antitermination activity *in vivo* to the same extent as full length transcripts in *trans* and within the section of the leader transcript residues 36-41 and 54-62 are independently required for the antitermination reaction *in vivo*.

1.7.8 Translation Coupling Exists Between the Regulatory Genes of the Amidase Operon of Pseudomonas aeruginosa.

For the amidase operon to operate efficiently regulation must be flexible enough to sense appropriate levels of inducing ligands for induction to occur and also be capable of producing large amounts of the activator component for mRNA binding.

A fixed stoichiometry of 3.3:1 is believed to exist between AmiC and AmiR under all growth conditions and this is responsible for the systems induction profile. Leaky operon expression under non-inducing growth conditions allows for sufficient amidase production to hydrolyse small amounts of internally generated acetamide and small amounts of AmiC and AmiR are produced to allow for operon induction under appropriate conditions. Upon induction, large amounts of amidase are produced, and lesser amounts of AmiR [for RNA binding], and AmiC [which is required to halt operon expression upon exhaustion of the ligand] [Norman *et al.*, 2000].

Since both *amiC* and *amiR* lie within the operon polycistronic transcript their translation will be dependent upon the strength and accessibility of their respective

Translation Initiation Regions [TIRs] and it has been suggested that a translation coupling system operates to maintain the AmiC:AmiR stoichiometry [Norman *et al.*, 2000]. The start codon of *amiR* and stop codon of *amiC* overlap - **AUGA** - and *amiC* has a good TIR which is ribosome accessible, whilst *amiR* has a poor TIR which is sequestered into a stem/loop structure. It is thought that AmiR cannot be produced independently - only after AmiC synthesis - since it has a poor TIR which is sequestered into a region of secondary structure. Thus AmiR is only synthesised after AmiC has been produced, and it occurs inefficiently by the ribosome back-tracking and re-initiating at the AmiR start codon about 1/3 of the times.

1.8 Project Aims

Macromolecular assemblies play essential roles in biological processes, a role which is crucial to the regulation of gene expression. Within these complexes, the interactions that occur are selective in forming unique structures. One such assembly is the complex formed between AmiR and the leader RNA of the amidase operon, in which binding to the RNA ligand by AmiR, results in transcription antitermination and subsequent read-through to the downstream genes by RNAP. The nature of the AmiR/RNA complex formation is not known, and it was the overall aim of this project to elucidate the mechanism of antitermination by AmiR.

2 Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

The bacterial strains used in this study are listed in Table 2.1 and Table 2.2.

Table 2.1. E. coli strains used in this study

Strains	*Genotype + Relevant Features*	*Reference/Source*
BL21 [DE3]	*E. coli B, F-, dcm, ompT, hsdS(rB-mB-), gall(DE3).* Contains T7 RNAP gene under control of the *lac*UV5 promoter, integrated into the chromosome (DE3).	Dubendorff, and Studier, [1991].
BL21 [DE3] pLysS	*E. coli B, F-, dcm, ompT, hsdS(rB-mB-), gall(DE3) pLysS.* High stringency expression host giving tight regulation of T7 RNAP expression.	Dubendorff, and Studier, [1991].
BMH 71 – 18	*Thi, supE, Δ[lac-proAB], mutS::Tn10, [F', proAB, lacIqZΔM15]*	Hanahan,[1983] [Promega]
DH1	*F-, endA1, gyrA96, thi-1, hsdR17, supE44, relA1.*	[Stratagene]
DH5-α	*F-,ϕ 80d lacZΔ M15, Δ (lacZYA-argF),U169*	[Promega]
JM109	*lacIq, recA-. e14-(mcrA-), recA, endA, gyrA96, thi-1, hsdR17(rK-mK+), supE44, relA, l-, Δ (lac-proAB), [F'traD36, proAB, lacIqZΔ M15]*	Hanahan, [1983] [Promega]
MAH1	*F-, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, ::pami/Leader.*	This study
SM10	*Thi, thr, leu, tonA, lacY, supE, recA::RP4-2-TC::mu, km*	Miller and Mekalanos [1988]
XL1-BluInt	*recA, endA, gyrA96, thi-1, hsdR17(rK-,mK+), supE44, relA1, l-, lac-, [F' proAB, lacIqZΔ M15, Tn10(Tetr)], ::pami lacZ*	[Stratagene]

Table 2.2. P. aeruginosa strains used in this study

Strains	*Genotype [Phenotype]*	*Reference/Source*
PAC1	*amiEBCRS* [inducible]	Kelly and Clarke [1962], [Drew, R]
PAC452	*amiΔ161* [Ami deletion]	Day *et al.*, [1975], [Drew, R]

2.1.2 Plasmids

Plasmids used in this study are listed in Table 2.3 and new plasmids constructed are listed in Table 2.4.

Table 2.3. Plasmids used in this study

Plasmid	*Size [bp]*	*Markers*	*Phenotype*	*Reference/Source*
pETDuet-1	5420	Amp	Protein Expression Vector	Novy *et al.*, [2002]
pGEM4Z	2746	Amp	Cloning Vector	Yanisch-Perron *et al.*, [1985]
pGEX6P1	4984	Amp	Protein Expression Vector	Smith and Johnson [1988]
pJB950	9101	Amp	*ami operon* expression vector	Clarke *et al.*, [1981]
pRAN1	4250	Amp	*amiC amiR* expression vector	Norman *et al.*, [2000]
pSW24	4286	Amp	*amiR* expression vector	Wilson *et al.*, [1996]
pSW35	4577	Sm	*amiR* expression vector	Wilson *et al.*, [1996]
pTM1	6879	Cm	*amiE/Leader* expression vector	Wilson *et al.*, [1996]
pUC18NotI^{Km}	2686	Amp	Cloning Vector	Herrero *et al.*, [1990]
pUTminiTn5^{Km}	7440	Km	Transposon Vector	Herrero *et al.*, [1990]
pUC19	2686	Amp	Cloning Vector	Yanisch-Perron *et al.*, [1985]

Table 2.4. Plasmids constructed for this study

Plasmid	*Parental Plasmid*	*Size [bp]*	*Marker*	*Genotype*
pAmiR.A152G	pSW24	4286	Amp	*amiR* A152G mutation
pAmiR.A152V	pSW24	4286	Amp	*amiR* A152V mutation
pAmR.A167G	pSW24	4286	Amp	*amiR* A167G mutation
pAmiR.A168A	pSW24	4286	Amp	*amiR* H168A mutation
pAmiR.H168R	pSW24	4286	Amp	*amiR* H168R mutation
pAmiR.H168E	pSW24	4286	Amp	*amiR* H168E mutation
pAmiR901.P20Q	pMH901	8331	Amp	*amiE* expression vector *amiR* P20Q mutation,
pAmiR901.V154E	pMH901	8331	Amp	*amiE* expression vector *amiR* V154 E mutation
pAmiR901.R179H	pMH901	8331	Amp	*amiE* expression vector *amiR* R179H mutation
pMH1	pGEM4Z	5100	Amp	*ami*-Leader *amiE* gene WT sequence
pMH2	pGEM4Z	5100	Amp	*ami*-Leader mutation
pMH3	pGEM4Z	4982	Amp	*ami*-Leader 118bp deletion
pMH31	pGEM4Z	4982	Amp	*ami*-Leader deletion
pMH4	pGEM4Z	4982	Amp	*ami*-Leader mutation
pMH41	pGEM4Z	4964	Amp	*ami*-Leader deletion
pMH5	pGEM4Z	4982	Amp	*ami*-Leader mutation
pMH51	pGEM4Z	4945	Amp	*ami*-Leader deletion
pMH6	pGEM4Z	4982	Amp	*ami*-Leader mutation
pMH61	pGEM4Z	4934	Amp	*ami*-Leader mutation
pMH8	pGEX6P-1	5131	Amp	ANTAR protein expression vector
pMH10	pUC18*Not*I	5106	Amp	*amiE*-Leader sub-cloning vector

Table 2.4. Plasmids constructed for this study continued

pMH11	pUTminiTn5^{Km}	9985	Km	*amiE transposon vector*
pMH90	pETDuet-1	7794	Amp	*ami*-Leader and *amiE* gene in MCS1
pMH91	pETDuet-1	5926	Amp	*amiR* gene in MSC2
pMH901	pETDuet-1	8331	Amp	*ami*-Leader and *amiE* gene in MCS1 +*amiR* gene in MSC2
pMH902	pETDuet-1	7893	Amp	*Ami* leader +*amiE* gene MCS1, *antar* MCS2
pMH92	pETDuet-1	5482	Amp	*antar* MCS2

2.1.3 Growth Media for Bacterial Strains

2.1.3.1 LB Broth [Luria-Bertani]

Bacto tryptone 10.0 g/L

Bacto yeast extract 5.0 g/L

NaCl [molecular biology grade] 10.0 g/L

Adjust the above to pH 7.0 with NaOH

Media was prepared by dissolving the components in dH_2O and sterilized by autoclaving.

2.1.3.2 LB Agar

To components of LB medium above, add 15.0 g/L of molecular biology grade agar.

2.1.3.3 M9 Minimal Medium Agar Plates

M9 salts were made up in 1L of distilled water.

$Na_2HPO4.7H_2O$ 64.0 g/L

KH_2PO4 15.0 g/L

NaCl 2.5 g/L

NH_4Cl 5.0 g/L

The solution was divided into 200 ml aliquots and autoclaved. The following was then added to 750 ml of distilled water,

200 ml M9 salts,

2 ml 1M $MgSO_4$,

1.1 ml 1M $CaCl_2$

1 ml Trace Elements

1 ml 0.01% Thiamine

15.0 g Agar.

The solution was autoclaved and allowed to cool to 60 C before the addition of filter sterilised acetamide [10 ml 30% acetamide] and appropriate antibiotics.

2.1.3.4 2 X TY Broth

Bactotryptone 16.0 g/L

Bacto yeast extract 10.0 g/L

NaCl (molecular biology grade) 5.0 g/L

Adjust the above to pH 7.0 with NaOH

2.1.3.5 2 X TY Agar

To components of 2X Ty medium above add 15.0 g/L agar.

2.1.4 Maintenance of Bacterial Strains and Antibiotics

E. coli strains were maintained as 20% glycerols stocks at –70 C. These stocks were prepared by washing 7 ml of 20% (v/v) sterile glycerol over an agar plate and resuspending the cells by repeated pipetting with a 10 ml Sterilin pipette. The resulting homogenous mixture was collected in a 5 ml capped Falcon tube. Overnight liquid culturing was achieved by inoculating 50 µl of this stock into 5 ml broth, whilst streaking from a sterile wire loop was sufficient for growth on agar plates.

Plasmid-harbouring bacteria were growth in the presence of their appropriate selective antibiotic to a final concentration of; Ampicillin [Amp] sodium salt [100μg/ml] for Ampicillin resistance, Chloramphenicol [Cm] [20μg/ml] for chloramphenicol resistance, Streptomycin [Sm] [80μg/ml] for Streptomycin resistance and Tetracycline [Tet] hydrochloride [20μg/ml] for Tetracycline resistance.

2.1.5 Buffers

2.1.5.1 Loading buffer for DNA Agarose gels

0.25% w/v xylene cyanol

0.25% w/v bromophenol blue

30% w/v glycerol

- in 10X TBE buffer

2.1.5.2 10X TBE buffer

TRISMA base 162 g/L

Boric acid 27.5 g/L

EDTA 9.25 g/L

2.1.5.3 DEPC-treated water

To 1 litre of deionised water, add 1ml of DEPC [Sigma] [0.1% final concentration]. Incubate overnight in the fume hood. Sterilise by autoclaving for a minimum of 20 mins at 1.0 bar [15 lb/psi].

2.1.5.4 IPTG [Isopropyl- beta-D-thiogalactopyranoside]

IPTG [Sigma] was obtained as a 1.0 g desiccated salt. 5ml of sterile DEPC treated H_2O was added to give a 200 mg/mL stock solution. IPTG was used at final concentration of 1.0 mM.

2.1.5.5 X-gal [5-bromo-4-chloro-3-indoyl-beta-D-galactopyranoside]

X-gal [Sigma] was purchased as a desiccated salt preparation and dissolved in dimethyl-fluoride [DMF]. X-gal was used at a final concentration of 50 μg/ml.

2.1.5.6 SSC, 20X

3M NaCl

0.3M sodium citrate, pH 7.2

2.1.5.7 TAE 50X buffer

TRISMA base 242 g/L, to this was added;

57.1 ml glacial acetic acid and

100 ml of 0.5 M EDTA, disodium salt, pH 8.0,

2.1.5.8 TE buffer

10 mM Tris-HCl, pH 8.0

1mM EDTA, disodium salt

2.1.5.9 Phosphate Buffered Saline

KCl 0.2g/L

NaCl 8.0 g/L

KH_2PO4 0.2 g/L

NaHPO4 1.15 g/L

2.2 Methods

2.2.1 Microbiological Methods

2.2.1.1 Preparation of competent E. coli cells.

5ml of LB Broth was inoculated with a bacterial colony from an agar plate or with 20µl of glycerol stock and the culture grown at 37 C on a reciprocating shaker until they reached an optimum density of [A650nm =0.6].

Competent *E. coli* cells were then prepared according to a calcium chloride technique as described in Maniatis *et al.*, [1989]. The competent cells were resuspended in ice cold 0.1M $CaCl_2$ plus 15% glycerol and stored in 0.1 ml aliquots at -70°C for up to 2 months.

2.2.1.2 E. coli Transformation

1ng-1µg of plasmid DNA was mixed with an aliquot of competent *E. coli* cells prepared as described above. Samples were iced for 30 min followed by a heat shock step at 42 C in a water bath for 1 min. Samples were plunged into ice and left to chill for 10 min. 1 ml of SOB medium or 2X TY medium was added, and the tubes shaken at 37 C for 1 hour. The cells were then pelleted and resuspended in 200 ml of medium containing appropriate antibiotics. The cells were then ready to be plated out.

2.2.1.3 Amidase Assays

Amidase was assayed in whole cells by the Transferase assay of Brammar and Clarke [1964] using acetamide as a substrate.

A solution of mixed substrates was freshly prepared using two volumes of 0.1M Tris HCL pH 7.2, one volume of 0.4M acetamide [freshly prepared] and one volume of 2M neutral hydroxylamine [10mL of 5M hydroxylamine hydrochloride neutralised to pH 7.2 with 10N NaOH and made up to 25mL with distilled water] and kept on ice. To 0.9 ml of prewarmed [37 C] mixed substrate solution, 0.1 ml of suitably diluted bacterial culture was added and the mixture incubated for 10 minutes. The reaction was stopped by the addition of 2 ml ferric chloride/HCL solution [100mL 60% ferric chloride and 57 ml of 12N HCL made up to 1L with distilled water]. Reactions were vortexed to dispense nitrogen bubbles and then read at 500nm using a spectrophotometer against a reagent blank. The absorbance was related to the amount of acetohydroxamate formed in the reaction by comparison with a standard curve [Brown *et al.*, 1969]. An absorbance of 1.0 at OD_{500} corresponds to 3.5 μmols acetohydroxamate formed/ml.

Amidase activity levels are presented as the mean values of duplicate assays carried out on at least three separate occasions. One unit represents 1.0 μmol of acetohydroxamate formed per min. Specific activities presented are units/mg dry weight bacteria. The following equation was used in the calculation of amidase specific activities:

$$3.5 \times [A500 - (A670 \times 0.08)] / [A670 \times 0.56]$$

2.2.1.4 96-well Plate Amidase Assay

Amidase assays carried out in 96-well plates were a modified version to allow reaction to occur in wells with a maximum volume of 500μL, therefore reactions were scaled down [x 10].

2.2.1.5 Construction of MAH1

To identify *amiR* mutants, an *E. coli* DH1 test strain was constructed [MAH1] with the *ami* promoter/leader and *amiE* inserted randomly into the chromosome by transposition. SM10 was the donor strain carrying pMH11, which has the *ami* sequences and a Km^R gene between Tn5 ends. Overnight LB cultures of SM10, pMH11 and DH1 were used to inoculate 5 ml of fresh medium and the strains grown for 180 min. 100 µl of each strain was spread onto a dried LB agar plate and incubated for 4 hours at 37C. The cells were then lifted after the addition of 3 ml dilution buffer, washed once with dilution buffer, and 100 µl of 0 fold, 10 fold and 100 fold dilutions were spread onto M9 minimal agar plates containing 20 % glucose, 0.01 % Thiamine and Kanamycin and incubated at 37 C.

2.2.2 Molecular Techniques

2.2.2.1 Isolation of Plasmid DNA

Plasmid DNA was routinely isolated and purified from 5ml or 100ml overnight *E. coli* cultures with the Qiagen QIAPrep Spin MiniPrep or MidiPrep Kits respectively for medium to large-scale plasmid DNA isolation, according to the methods described by the manufacturer.

2.2.2.2 Agarose gel electrophoresis of DNA

DNA fragments between 0.5 and 25kb were routinely separated by agarose gel electrophoresis using TBE or TAE buffers. Ethidium bromide was added to the agarose at a final concentration of 0.5µg/ml, before being poured onto a sealed platform with the appropriate gel combs. DNA samples were electrophoresed [at 1 to 10 V/cm of agarose gel], alongside DNA molecular weight markers. Gels were subsequently viewed using short wave UV transilluminator and images were digitally captured.

2.2.2.3 Isolation of DNA Fragments from Agarose Gels

PCR fragments and digested linear DNA, were extracted from Agarose gels using the Qiagen Gel Extraction Kit. DNA bands were visualised on a Flowgen Transilluminator at 365 nm to minimise nicking of the nucleic acid backbone. DNA was excised from the gels using fresh, clean scalpels and the protocol followed according to the manufacturers instructions.

2.2.2.4 Quantification of DNA

DNA samples were quantified as described in Maniatis *et al.*, 1989. Spectophotometric readings of the samples were taken at a wavelength of 260nm. An optical density [OD] of 1 corresponds to approximately 50 μg mL-1 for double stranded DNA.

2.2.2.5 Enzymatic Manipulation of DNA

Enzymatic modifications of DNA such as restriction enzyme digestions and ligations were usually carried out under conditions of DNA, buffer and enzyme concentrations, specific temperature and incubation periods as recommended by the suppliers. All enzymes used were from NEB unless otherwise stated.

2.2.2.6 Restriction Enzyme Digestion

Typically, 0.2 to 1.0 μg plasmid DNA [in sterile H_2O, or TE Buffer] was digested with 1μL [1 to 10 units] of restriction endonuclease and 2 μl of the recommended 10 x NEB Buffer and 100 μg/ml BSA if required. The mixtures were then adjusted to a total reaction volume of 20 μl with sterile H_2O and incubated at the appropriate temperature for two hours to allow complete digestion. Inactivation of the enzyme activity when required was achieved by incubation at the recommended heat inactivation temperature. Where this was not possible, restriction enzyme digests were purified and enzymes inactivated using a Qiagen PCR Purification Kit.

2.2.2.7 Dephosphorylation of 5' linear DNA ends

Removal of 5' phosphate groups from linearised DNA molecules was achieved by adding 0.5 µl Calf Intestinal Alkaline Phosphatase [CIP] to the digested DNA and incubated at 37 C for 30 minutes. CIP activity was inactivated by addition of EDTA [pH 8.0] to a final concentration of 10mM and incubated for 10 minutes at 70 C.

2.2.2.8 Creation of 'blunt ended' linear DNA

When necessary for further ligation, the 5' or 3' recessed termini of linear DNA were filled in using either DNA Polymerase I Large Fragment [Klenow] from *E. coli* or bacteriophage T4 DNA Polymerase. DNA was quantified approximately by comparisons to known DNA size markers after gel electrophoresis and overhangs filled in as follows:

2.2.2.9 DNA Polymerase I Large Fragments [Klenow]

1 uL of enzyme was added to an 2-10 µg DNA and the mixture supplemented with 10 µM dNTPs and 1/10th the final volume of 10 x *EcoPol* Buffer. The reaction mixture was incubated at 25 C for 15 minutes. The enzyme was inactivated by the addition of EDTA to a final concentration of 20 mM and incubated at 75 C for 20 minutes.

2.2.2.10 T4 DNA Polymerase

1µL of enzyme was added to 2-10 µg and supplemented with 10 µM dNTPs, 1/10th the final volume of 10 x T4 DNA Polymerase Buffer and BSA to a final concentration of 50µg/mL. Reactions were incubated at 12 C for 20 minutes then heat inactivated for 10 minutes at 75 C.

2.2.2.11 Polymerase Chain Reaction [PCR] Amplification of DNA

PCR was used for DNA amplification and site-directed mutagenesis.

A 50 µl preparative PCR reaction contained 5 µl of 10X Expand™ high fidelity buffer with 15 mM $MgCl_2$, 20 mM Tris-HCL - pH 7.5, 100 mM KCl, 1 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.5% Tween 20 (v/v), 0.5% Nonident® p40 (v/v), 50% glycerol (v/v), 100 µl dNTP mix (100 µM of each of dATP, dCTP, dTTP, dGTP)], 100 pmol of each primer, and 1.14 unit Expand™ HF polymerase [Roche, Germany].

Table 2.5. PCR conditions

Number of cycles	*Conditions*
1x	Denature template 2 min. at 94°C
35-40x	Denaturation at 94°C for 30 sec Annealing usually at 45-65°C* for 30 sec Elongation at 72 or 68°C** for 45sec-8min Table 2.6.
1x	Final elongation at 72°C for min

* Annealing temperature depends on the melting temperature of the primers used.

** Elongation temperature depends on the length of amplification product: 72°C was used for amplification up to 3.0 kb; 68°C was used for amplification >3.0 kb.

Table 2.6. PCR Elongation Time

Elongation time	45 sec	1 min	2 min	4 min	8 min
PCR fragment length (kb)	<0.75	1.5	3.0	6.0	10.0

2.2.2.12 Ligation of DNA Fragments

Ligations were carried out in a final volume of 20µl containing 2µl of the manufacturers' recommended 10 x ligation buffer and 1µl (10 units) T4 DNA ligase at 14 C overnight.

2.2.2.13 Restriction analysis of plasmids for detecting positive clones

Plasmid clones were initially screened by restriction digest and electrophoresis. This allowed identification of clones containing the correct insert. Recombinants containing inserts of the correct size were confirmed to have the correct insert by sequencing.

2.2.2.14 Site Directed Mutagenesis of DNA

Site directed mutagenesis was carried out using the QuickChange® Site-Directed Mutagenesis Kit [Stratagene] and the GeneEditor™ in vitro Site-Directed Mutagenesis System [Promega]. Reactions were carried out according to the manufacturers instructions.

The Quikchange™ site-directed mutagenesis method utilises a double-stranded vector and two synthetic oligonucleotide primers [containing the desired mutation], each complementary to opposite strands of the vector. The mutant oligo is extended by PCR and a mutated plasmid containing the staggered nicks is generated. Following temperature cycling the product is treated with *Dpn*I [tagged sequence: 5'-G^{m6}ATC-3'] which is specific for methylated and hemi-methylated DNA. The original DNA template is *dam* methylated and susceptible to *Dpn*I digestion while the PCR generated mutant plasmid DNA is unmethylated and resistant to digestion leading to a high mutation efficiency. After PCR extension and *Dpn*I digestion, mutated plasmid DNA was transformed into JM109 competent cells as described in Section [2.2.1.2]. Plasmid DNA was isolated from single JM109 transformants, and checked by restriction digest and sequencing.

2.2.2.15 Random Mutagenesis I: Hydroxylamine Mutagenesis

A mixed solution was freshly prepared containing 1ml of 5M hydroxylamine hydrochloride [1M], 500 µl 500 mM sodium pyrophosphate, 500 µl 1M NaCl, 100 µl 100 nM EDTA and 3ml H_2O to make up a 5 ml volume.

10 µg of target DNA was added to 500 µl of hydroxylamine solution, vortex mixed and incubated at 75 C. At time intervals: 0, 5, 10, 15, 30, 60, and 90 minutes, 100 µl of the reaction mixture was removed and iced. The hydroxylamine was removed using QIAgen PCR purification columns, and the DNA eluted with 30 µl EB Buffer.

200 µl of competent DH1, pTM1 cells were transformed with 1µl of each of the time points of DNA. Dilutions were plated onto a] MM + acetamide + Cm +Amp and b] Amp LB plates to calculate the transformation efficiency. A time point was chosen which had a mutation rate closest to 5% i.e. the average colony number is 95% that of the timepoint 0 transformation.

2.2.2.16 Random Mutagenesis II: Error prone PCR

Random mutagenesis was carried out using the GeneMorph® II Random Mutagenesis Kit [Stratagene] according to the instructions. The rate of Random Mutagenesis was optimised by adjusting template concentrations and PCR cycle number.

2.2.3 Protein Expression and Purification Methods for ANTAR

2.2.3.1 Reagents, buffers for GST-Sepharose affinity chromatography

All buffers for GST-Sepharose affinity chromatography were stored at 4 C prior to use.

2.2.3.2 Lysis buffer

10 % YPER [Pierce Chemicals]

50 mM NaCl

20 mM Tris-HCl [pH adjusted to 1 unit out of the pI of the desired protein].

2 mM DTT

2 mM EDTA

Protease Inhibitor Cocktail [2 tablets/ 50 mL] [Roche]

DNase [10 U/ 50 mL] [Turbo Dnase (Ambion)].

2.2.3.3 Buffer A [Wash buffer]

1 M NaCl

20 mM Tris-HCl [pH 7.4].

2 mM DTT

2 mM EDTA

2.2.3.4 Buffer B [GST -Binding buffer and GST-cleavage buffer]

50 mM NaCl

20 mM Tris-HCl [pH 7.4].

2 mM DTT

2 mM EDTA

2.2.3.5 Buffer C [Gel filtration chromatography buffer]

300 mM NaCl

20 mM Tris-HCl [pH adjusted to 1 unit out of the pI of the desired protein].

2 mM DTT

2 mM EDTA

Buffers for gel filtration were filtered through a 0.1 micron membrane filter, degassed extensively and chilled to 4°C before use.

2.2.3.6 Buffer D [ANTAR Dilution Buffer]

300 mM NaCl

20 mM Tris-HCl

2 mM DTT

2 mM EDTA

pH 7.4

2.2.3.7 SDS polyacrylamide gel electrophoresis

Proteins can be separated according to their size using a discontinuous electrophoresis system. Slab gels were cast in the BioRad Mini-Protean-1 system. Gel solutions are listed below. This system consists of a stacking gel, containing 3% acrylamide and a resolving gel containing the appropriate percentage of acrylamide for efficient separation of protein or fragments of interest [ranging form 7-15% acrylamide]. Both gels were routinely cast using a 40% acrylamide:bis-acrylamide stock and polymerised upon addition of 0.05% [v/v] TEMED and 0.7% [w/v] APS. Prior to loading, protein samples were boiled for 10 mins in sample buffer. Samples were then loaded into wells formed within the stacking gel, and electrophoresis proceeded at a constant current of 30 mA until the appropriate level of separation had occurred. Molecular standard markers were run beside the samples.

2.2.3.8 SDS-PAGE Solutions

30% Acrylamide stock [BDH] 29.2% acrylamide and

0.8% Bis-acrylamide

APS 10% [w/v] solution stored at 4 c

Stacking Gel

125mM Tris-HCL [pH6.8]; 0.1 SDS, 4% polyacrylamide

Resolving Gel

500mM Tris-HCL [pH8.8]; 0.1 SDS 7% polyacrylamide

Protein Sample Buffer

0.125M Tris-HCL [pH6.8] 10%[v/v] glycerol, 5% [v/v]

β-mercaptoethanol; 2% SDS and 0.015% [v/v] bromophenol blue.

2.2.3.9 Growth conditions and IPTG Induction for protein expression from pGEX6P1

BL21 [DE3] pLysS cells that had been transformed with recombinant pGEX6P1 plasmids were spread onto LB-agar-plates containing Amp. Plates were incubated overnight at 37 C. On the following day, single colonies were used to prepare confluent streaked plates, again by overnight incubation at 37 C. 10 ml of LB Amp was pipetted onto each plate. A glass spreader was used to lift the colonies gently off the plates into the medium, which was then used to inoculate a single conical flask containing 800-900 ml of LB-medium/100mg/ml ampicillin. Flasks were incubated in a shaking incubator [220 rpm] at 37 C. The growth of the bacteria was monitored until O.D. 600nm =1.0

The temperature of the incubator was lowered to 25 C and flasks were allowed to equilibrate at this temperature before addition of 1mM IPTG to induce the recombinant fusion protein. The flasks were then shaken at 25 C for 5 hours, after which the cells were recovered by centrifugation at 3000 x g at 4 C. The cell

pellets were stored at –20 C until a convenient time to harvest the recombinant proteins.

2.2.3.10 Lysis of cells by sonication.

Bacterial cell pellets stored at –20 C were allowed to thaw on ice, then they were resuspended in lysis buffer [50mM NaCl, 20mM Tris-HCl, 2mM DTT, 2mM EDTA, protease inhibitors (Complete EDTA Free, Protease Inhibitor Cocktail Tablets Roche {Applied Sciences} 1 tablet/50 mL)]. The pH of the buffer was adjusted to within 1 unit out of the estimated pI of the recombinant protein [for example, the pI of the recombinant protein was 6, the pH of the buffer was adjusted to 7, likewise if the pI of the protein was 9, then the pH of the buffer was adjusted to 8].

Cell suspensions were sonicated on ice with 6 pulses of 80 watts for 30 sec, with 2 minutes rest in between. Between each pulse, the suspensions were gently stirred with a clean spatula. The sonicated cells were centrifuged at 18 000 x g for 1 hour at 4 C, and the supernatant containing the soluble protein fraction was aspirated into a fresh flask.

2.2.3.11 Binding of soluble protein fraction onto GST-sepharose beads

A KX-16 column [Pharmacia] was packed with approximately 2 ml of GST-sepharose beads. The column was then washed with 5 column volumes of milli-Q water, equilibrated with 5 column volumes of Buffer A [2.2.3.3]. The supernatant from the lysed cells were then loaded onto the GST-sepharose column using a P1 pump. The column was then connected to a FPLC instrument and washed with buffer B [2.2.3.4] at a flow-rate of 1.5 ml/min. The O.D. of the exudate was monitored at 280 nm, and once the trace had reached baseline, the column was equilibrated once again with Buffer A.

2.2.3.12 Cleavage of recombinant fusion protein with PreScissionTM protease.

The GST-sepharose beads were transferred from the KX-16 column into a 50 ml Falcon tube. Two column volumes of Buffer A [without protease inhibitors] were added to the beads. Two units of PreScission™ protease [Amersham-biosciences] are needed for 100mg of bound GST-fusion protein, thus approximately 100 units of the enzyme were added to 2 ml of GST-beads and left on a rolling-incubator at 4 C overnight. The PreScission™ protease is a genetically engineered fusion protein consisting of the human rhinovirus 3C protease and GST and specifically recognises the amino acid sequence Leu-Glu-Val-Leu-Phe-Gln†Gly-Pro. Cleavage occurs between Gln and Gly. The protease is itself fused to GST and therefore easily removed from the cleavage reaction using glutathione sepharose.

2.2.4 Gel Filtration Chromatography Methods

2.2.4.1 Concentration of sample prior to gel filtration

The eluate from GST-sepharose beads was concentrated using 2 x 20ml centricon [Amicon] concentrators. Centricon tubes containing the samples were placed in a table-top centrifuge and spun at 3500 x g at 4 C. The concentrated samples were inspected visually every 30mins to see if any protein aggregation [visualised as white flakes] or precipitation was occurring. The sample volume was brought down to 4.5 ml, suitable for injection into a 5-ml loop of an AKTA prime.

2.2.4.2 Gel filtration running conditions

Gel Filtration was carried out using a 300 ml S200 column. Prior to use, the column was equilibrated with 1 column volume of Buffer C. The S200 column was connected to an AKTA -prime [Amersham Biosciences]. The U.V. detector was set to monitor the column eluate at 280 nm. The flow-rate was set to 1.5 ml/ min and 6-ml fractions were collected.

2.2.5 *Dynamic Light Scattering*

2.2.5.1 Sample preparation for Dynamic Light Scattering

Prior to analysis by DLS, the cuvettes were washed with clean millipore water and dried. 12 ml of water filtered through a 0.1 mm filter [Whatman, UK] was dispensed into the cuvette and the counts measured at 17 C. The average count rate of pure clean water should be between 23 - 25 counts [if the counts were fluctuating beyond these values, it was an indication that the cuvette needed more thorough cleaning as foreign particles were affecting the background measurements].

The protein sample was spun in a micro-centrifuge at > 10 000 x g to sediment any precipitate that may interfere with the DLS measurements. The protein was then filtered through a 0.1mm filter [Whatman, UK] and 12 ml of the protein solution was dispensed into a clean cuvette for measurement.

2.2.6 *Mass Spectrometry of ANTAR*

2.2.6.1 Sample Preparation for Tryptic Digestion of Protein Samples directly in Polyacrylamide Gel Slices for Mass Spectrometry

Sample bands were excised from the gel and cut into several smaller pieces. These were destained overnight in a volume of 500µl at room temperature. The following day, the destain was discarded and the gel pieces allowed to dehydrate in 200 µl acetonitrile. The acetonitrile was then discarded and any residual acetonitrile removed in a SpeedVac.

The gel pieces were then reduced in 30-50 µl 10 mM DTT for 30 min at room temperature. Samples were then alkylated in 30-50 µl 50 mM iodoacetamide for 30 min room temperature, followed by washing with 100ul 100 mM ammonium bicarbonate for 10 min. Samples were then dehydrated in 200 µl acetonitrile for 5

minutes. The acetonitrile was then removed and the samples were rehydrated in 100µl 100mM ammonium bicarbonate for 10 min, followed by another dehydration step with acetonitrile as before. Samples were then dried in a SpeedVac. To prepare trypsin digestion; Promega trypsin was dissolved in 1 ml ice cold 50 mM ammonium bicarbonate [trypsin concentration = 20 ng/µl and kept on ice]. 50µl of the trypsin solution was added to the gel sample to cover the pieces and incubated for 5 to 10 min on ice. Excess trypsin solution was removed and 20 µl 50mM ammonium bicarbonate was added to the samples and allowed to react overnight at 37°C. Samples were then extracted with 30µl 100 mM ammonium bicarbonate. The mixture was vortexed then incubated for 10 min. Samples were then microfuged, and the supernatant transferred to a clean 0.5 ml microfuge tube. Peptides were then extracted by adding 30µl extraction solution, incubated for 10 minutes at 37°C, finally samples were then evaporated to 20µl.

2.2.7 RNA-Protein Interaction Studies

2.2.7.1 DNA template preparation for in vitro transcription

Template DNA was prepared by digesting 15µg of plasmid pMH41 with *Bst*II for 2 hours at 60 c.

2.2.7.2 In vitro transcription

In vitro transcribed ^{32}P labelled leader run off RNA transcripts containing the antiterminator sequence were produced using the MaxiScript® Protocol [Ambion] kit and pMH41 as a template. For each ^{32}P labelled *in vitro* transcription reaction, an unlabelled reaction was run alongside the reaction. Reactions were carried out according to the MaxiScript® Protocol except the $[\alpha]^{32}P$ labelled residue was ATP rather than labelled UTP as described in the instructions.

Reactions were assembled on ice and carried out in a final volume of 20µl containing the following constituents;

8.5 µl Nuclease Free H_2O, 2 µl DNA template [2 µg], 2 µl 10 x Transcription Buffer, 1 µl CTP [10 mM], 1 µl GTP [10 mM], 1 µl UTP [10 mM], 2.5 µl ^{32}P labelled ATP*, and 2 µl T7 RNA Polymerase Enzyme.

*For the cold [unlabelled] reactions, 2.5 µl ATP was substituted for ^{32}P ATP.

2.2.7.3 Removal of unincorporated ^{32}P ATP and DNA template

Following transcription, the DNA template was digested with 1U of RNase-free DNaseI. To remove unincorporated [α]^{32}P ATP's from the labelled *in vitro* transcribed RNA, a nuclease free G-25 Sephadex column was used [Amersham Biosciences]. The entire reaction mixture was applied to an equilibrated G-25 column and centrifuged at 735 x g for 2 minutes. RNA was recovered by precipitation with 1 volume of 7.5M NH_4OAc and 4 volumes of ethanol Samples were stored at -70 c for up to two weeks.

2.2.7.4 Electromobility Shift Assays

Probes at a concentration of 2500 cpm were used in 25 µl reactions containing the following;

21.37 µl EMSA Buffer [Buffer conditions are given in Chapter 5]

1µl BSA

1 µl H_2O

0.63 µl [2500 cpm] Probe

1 µl ANTAR or FE of undiluted or diluted protein sample.

Protein samples had been previously diluted to give a final concentration of 1 μg/μl [undiluted], 0.1 μg/μl [1:10 dilution] and 0.01 μg/μl [1:100]. Samples were incubated for 45 min at 37 C for ANTAR binding and 65 C for the F/E protein.

To 12.5 μl from each sample, 5μl of RNA loading buffer was used for loading onto a 4-20% PAGE Gradient TBE [Invitrogen / Novex, 4-20% TBE Gels, 1.0 mm wells] gel and run at 180V for 1 hour at room temperature in 1 x TBE [100 mM Tris/100 mM boric acid/5 mM EDTA, pH 7.0]. After drying, the RNA-Protein complexes were visualized on a PhosphorImager.

3 *Isolation and Analysis of Antitermination Defective AmiR Mutants*

3.1 *Site Directed Mutagenesis of the Conserved ANTAR Residues*

3.1.1 *Introduction*

The C-terminal domain of AmiR has been identified as an ANTAR domain [Shu *et al.*, 2002]. This domain is found in the transcription antiterminator proteins NasR of *Klebsiella oxytoca* and NasT of *Azotobacter vinelandii*, the putative antiterminator protein RV1626 of *Mycobacterium tuberculosis*, and a number of real and hypothetical proteins. In AmiR, this domain folds into a 3-helix bundle, and previous experiments have shown that truncated mutants of AmiR, lacking the last helix, or the last two helices [Δ 179-195, and Δ 162-195 respectively] lack transcription antitermination activity, even though they are stable proteins [O'Hara *et al.*, 1999]. The ANTAR domain is also found in the RV1626. The crystal structure of RV1626 was recently solved and structure analysis shows RV1626 to have an ANTAR domain and an overall similar structure to that of AmiR [Morth *et al.*, 2004] [Figures 1.9 and 3.1]. RV1626 in solution was shown to be a monomer [one molecule per asymmetric unit] and the ANTAR domain retains the overall fold as seen in AmiR. Interestingly, the major difference between AmiR and RV1626 is the presence of a kink in the middle region of the long hydrophobic coiled helix between the N - and CTDs of RV1626.

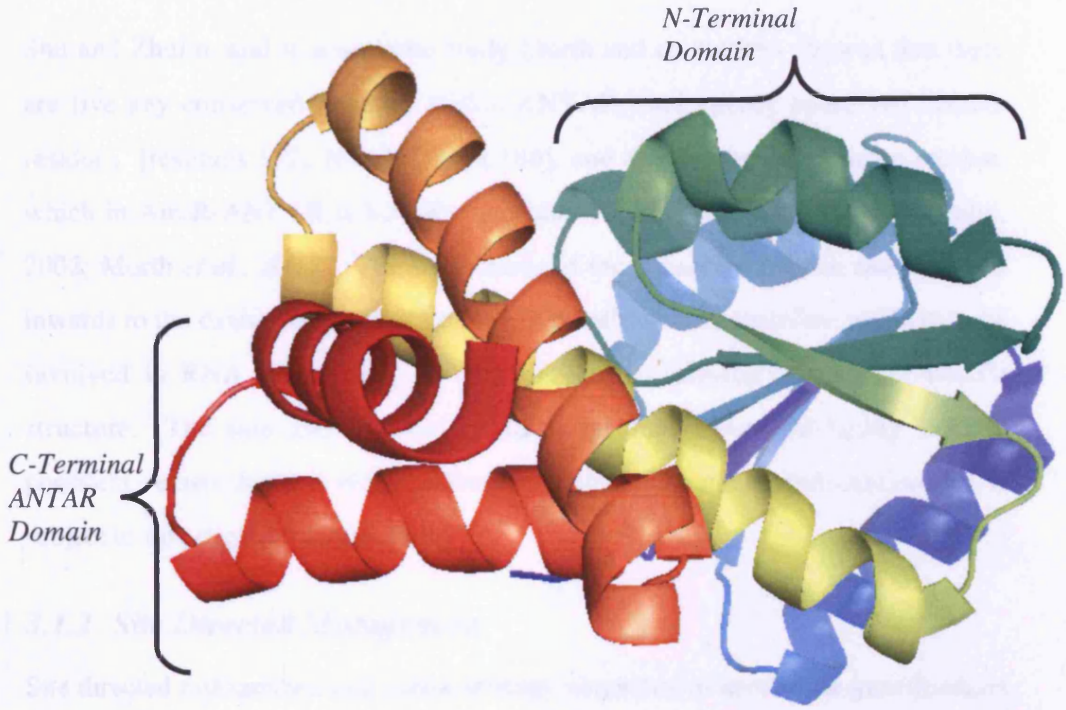


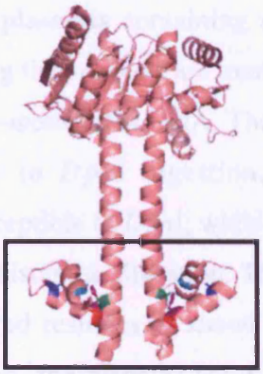
Figure 3.1. Structure of RV1626 from <u>Mycobacterium tuberculosis</u>.
The NTD shows structural homology to the NTD of AmiR and the CTD has a high structural homology to the ANTAR domain of AmiR [Morth et al., 2004]. Image created using PDB file 18SN using 3D Structure Viewer Pymol.

Additionally, RV1626 contains a phosphorylation site within the NTD not present in AmiR. The molecule in solution was found to be monomeric as it is in the crystal, but in solution it can undergo a conformational change triggered by changes in ionic strength, indicating that RV1626 is flexible and the crystal structure seen here is possibly one structural configuration out of many.

Shu and Zhulin, and in a separate study Morth and co-workers showed that there are five key conserved residues within ANTAR: four strictly conserved alanine residues [residues 152, 167, 175, and 186], and one conserved aromatic residue, which in AmiR-ANTAR is histidine [residue 168] [Figure 3.2] [Shu and Zhulin, 2002; Morth *et al.*, 2004]. The side chains of the conserved alanine residues face inwards to the cavity formed by the 3-helix bundle and are therefore unlikely to be involved in RNA interaction, but are thought to stabilise the 3-helix bundle structure. The side-chain of the histidine residue is outward facing and the possibility exists that it is required for RNA interaction and antitermination, and is subject to investigation in this study.

3.1.2 Site Directed Mutagenesis

Site directed mutagenesis is a useful strategy employed to investigate gene function and protein structure and function, ligand binding capacity, protein/protein interactions, and protein/nucleic acid interactions. There are several different methods in achieving such mutations in DNA. *In vitro* site directed mutagenesis, or oligonucleotide-directed mutagenesis involves hybridising a synthetic oligonucleotide complementary to the target DNA except for a mismatch near the centre. This mismatched region will contain the desired mutation. The synthetic oligonucleotide also serves as a primer for initiation of DNA synthesis *in vitro*. Subsequent extention of the oligonucleotide by DNA polymerase generates a double stranded DNA that is heteroduplex with respect to the desired mutation.



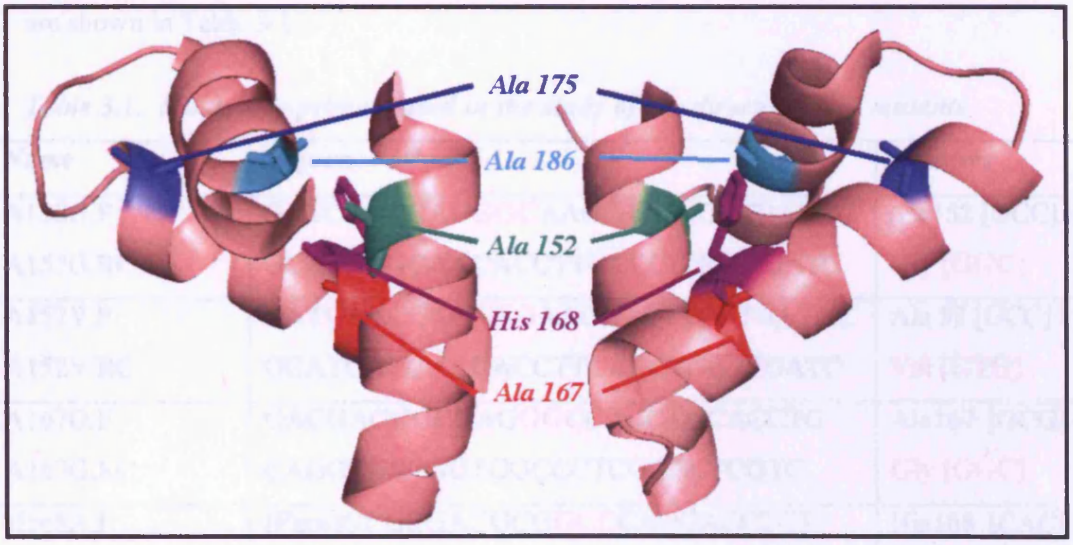


Figure 3.2. Conserved residues of the AmiR ANTAR domain.

Locations of the 5 strictly conserved residues within the ANTAR domain of the AmiR dimer. Alanine 152 [green], alanine 167 [red], alanine 175 [blue], alanine 186 [aqua] and histidine 168 [purple]. Image created from PDB file 1QO0 using 3D structure viewer Pymol.

The Quikchange™ site-directed mutagenesis method utilises a vector carrying the gene of interest and two synthetic oligonucleotide primers [containing the desired mutation], complementary to each strand of the vector. Mutant DNA is extended by PCR and two heteroduplex plasmids containing staggered nicks are generated. Following temperature cycling the products are treated with *Dpn*I, which is specific for *dam* methylated and hemi-methylated DNA. The original DNA template is *dam* methylated and susceptible to *Dpn*I digestion, heteroduplex molecules are hemimethylated and also susceptible to *Dpn*I, while PCR generated mutant plasmid DNA is unmethylated and resistant to digestion. This strategy was used to change the ANTAR domain conserved residues as shown in Table 3.1. Following PCR extension and *Dpn*I digestion, the plasmid DNA was transformed in XL1-Blue Supercompetent Cells. Plasmid DNA was isolated from single transformants, and checked by sequencing. The changes made and mutant oligonucleotide sequences are shown in Table 3.1.

Table 3.1. Mutagenic primers used in the study of site directed AmiR mutants

Name	*Sequence [5' –3']*	*Features*
A152G.F A152G.RC	GATCAACCAG**GGC**AAGGTGTTGCTGATGC GCATCAGCAACACCTTGCCCTGGTTGATC	Ala152 [GCC] to **Gly [GGC]**
A152V.F A152V.RC	GATCAACCAG**GTG**AAGGTGTTGCTGATGC GCATCAGCAACACCTTCACCTGGTTGATC	Ala 52 [GCC] to **Val [GTG]**
A167G.F A167G.RC	GACGAGCGCGAG**GGC**CACCAGCACCTG CAGGTGCTGGTGGCCCTCGCGCTCGTC	Ala167 [GCG] to **Gly [GGC]**
H168A.F	[Phos]GCGCGAGGCG**GCC**CAGCACCTGT CGCGGGAAGCG	His168 [CAC] to **Ala [GCC]**
H168R.F	[Phos]GGACGAGCGCGAGGCG**CGC**CAGC ACCTGTCGCGG	His168 [CAC] to **Arg [CGC]**
H168E.F	[Phos]GGACGAGCGCGAGGCG**GAG**CAGC ACCTGTCGCGGG	His168 [CAC] to **Glu [GAG]**

3.1.3 Effects of Mutations to Conserved Residues Found Within the ANTAR Domain

Amidase can be assayed in whole cells by the Transferase assay of Brammar and Clarke [1964] using acetamide as a substrate, as described in section [2.2.1.3]. The principle of the assay is based on the colour detection of acetohydroxamate produced in the reaction between acetamide and hydroxylamine. For this study, residues Ala-152, Ala-167 and His-168 were substituted for different amino acids and the effects of the changes analysed by an *in vivo* amidase assay. Plasmid pSW24 was chosen for the mutagenesis, which is a pUC19 derivative carrying a 1.5kb *Xho*I DNA fragment including the WT *amiR* gene expressed from the vector promoter. Functional analysis of the mutants was carried out by transforming the new plasmids into *E.coli*, pTM1, which carried *amiE* on a compatible pACYC184 derivative.

Amidase assays were performed on over-night cultures of original *E.coli*, pTM1 host, *E.coli* + pTM1+ pSW24 and *E.coli* + pTM1+ pSW24 mutants [Table 3.2]. Under these conditions, *E.coli*, pTM1 shows a specific activity of 1.0, representing low read-through of the upstream terminator – background amidase activity. In the additional presence of pSW24 [WT *amiR*], a SA of 37.0 is found – representing full AmiR-dependent antiterminated amidase activity.

3.1.3.1 Changes to Alanine 152

Two changes were made to residue A152, A152>G and A152>V, both increasing and decreasing the size of the R group. In both cases a decrease in amidase SA was measured. A152>G caused a 50% reduction and A152>V resulted in a 40% reduction.

3.1.3.2 Changes to Alanine 167

An A167>G mutation was identified and shown to almost completely abolish the antitermination activity of AmiR with activity reduced to 2.4.

3.1.3.3 Changes to Histidine 168

Histidine 168 is embedded in the penultimate helix with the R group pointing out into solution. The H168 mutations were made by project student [Bijal Amin] working under my direct supervision using the Promega Gene Editor System. Mutations H168>A, H168>R and H168>E were identified and characterised. The H168>A change caused a 69% reduction in amidase activity, and the activities for H168>R and H168>E indicated that these mutant AmiR molecules had lost all antitermination activity as only background activity was measurable.

Table 3.2. Specific Activities of AmiR ANTAR Domain Site-directed Mutants.

Plasmid	*Specific Activity*
pTM1	1.0
pTM1, pSW24	37.0
pTM1, pAmiR152A>G	18.6
pTM1, pAmiR152A>V	22.0
pTM1, pAmiR167A>G	2.4
pTM1, pAmiR168H>A	11.35
pTM1, pAmiR168H>R	1.0
pTM1, pAmiR168H>E	0.9

As would be expected, changes to highly conserved amino acid residues thought to be important in maintaining the overall structure of the CTD of AmiR led in all cases a reduction in antitermination *in vivo*. Replacement of Ala152 with either a

smaller G or larger V hydrophobic residue led to a reduction but not a complete loss of antitermination activity. The R group sits in a hydrophobic pocket made by the side chain of A147, A149, L155, A186 and L189 so the smaller G or larger V side chains would not be chemically out of place in this environment. The size changes however, may have altered the overall CTD structure.

3.2 Random Mutagenesis and Isolation of AmiR Mutants

3.2.1 Introduction

To aid the understanding of the mechanism by which AmiR functioned in the antitermination reaction, it was felt that the identification of critical residues was sufficiently important to undertake random *amiR* mutagenesis. The crystal structure of AmiR had shown the global structure in its inactive form. To gain more of an insight into the transformation of AmiR into its active state, a library of randomly generated mutants was created. These proteins were cloned into an expression system whereby their effects on antitermination could be measured. Previous experiments had generated some data, implicating regions within the ANTAR domain as well as with the NTD of AmiR, and it was therefore the aim of this study to isolate novel mutants with antitermination activities, hence shedding some light on important regions and structures of the AmiR protein required for antitermination.

3.2.2 Developing a Detection Method for the Random AmiR Mutants

3.2.2.1 Suicide Selection

It was initially hoped that a system whereby defective mutants were selected using a suicide selection method would host a suitable system of detecting mutants.

The basis for this was the finding that *E.coli* W3110 prototroph carrying pTM1 and pSW24 constitutively expressing high levels of amidase activity when grown on acetamide minimal medium were growth restricted. The reason for this is unclear but after several days incubation small faster growing colonies appear – surrounding the original cells – which show much reduced amidase activity. Under these growth conditions the original cells die or are very much growth restricted and mutants with much lower levels of amidase activity can be isolated. It was hoped that in this way new *amiR* mutants with reduced antitermination activity could be isolated and characterised.

To simplify the system for analysis of mutants, an *E.coli* DH1 derivative was constructed with a chromosomal *amiE* gene insertion. This chromosomal insertion was achieved by using a transposon vector to randomly insert this sequence into *E.coli* producing the strain MAH1 using a method developed by Herrero, and others [1990]. Derivatives of the Tn5 Transposon can be used to randomly insert DNA fragments cloned between the Tn5 IR ends in a plasmid vector into the *E.coli* chromosome [Herrero *et al.*, 1990]. For this experiment, the *amiE/Leader* gene fragment was gel purified and cloned into pUC18Not *Hin*dIII/*Eco*RI site following cleavage from pMH1 [Section 5.1.2.1.] using these enzymes to form the vector pMH10 [Figure 3.3 (a, b)]. *Not*I is a rarely cutting restriction enzyme [GC†GGCCGC] and there are no targets in the *ami* sequence. Cleavage with *Not*I generates the fragment with a *Not*I sticky end at each end and ligation into *Not*I cut pUTminiTn5km generated the donor strain pMH11 [Figure 3.3 (c, d)]. This has the cloned *amiE* fragment and Km^R gene inserted between Tn5 IR ends. Thus in the presence of Tn5 transposase the fragment can be randomly transposed into a new locations within the cell. The donor pMH11 plasmid carries the Tn5 transposase external to the transposable region, an R6K origin of replication and the plasmid RP4 *mob* site. Such R6K derivative plasmids only replicate in the presence of the *pir* activator protein – which is missing from this plasmid. However, the *pir* gene is present as a λ*pir* lysogen integrated into the host cell chromosome of strains CC118 and SM10. The ligation mix of the *Not*I *ami* fragment and pUTminiTn5 was transformed into competent SM10. In this strain pMH11 was stable and was maintained. In addition, SM10 carries an integrated RP4 plasmid, which can mobilise any *mob*+ plasmid to a new host. Thus mating this strain with other *E.coli* strains led to the mobilisation of the plasmid carrying the *amiE* Km^R gene into the recipients. Here it was unable to survive as the *pir* function is missing and selection of Km^R transconjugants generated those strains in which the *amiE* Km^R DNA had been transposed into the chromosome.

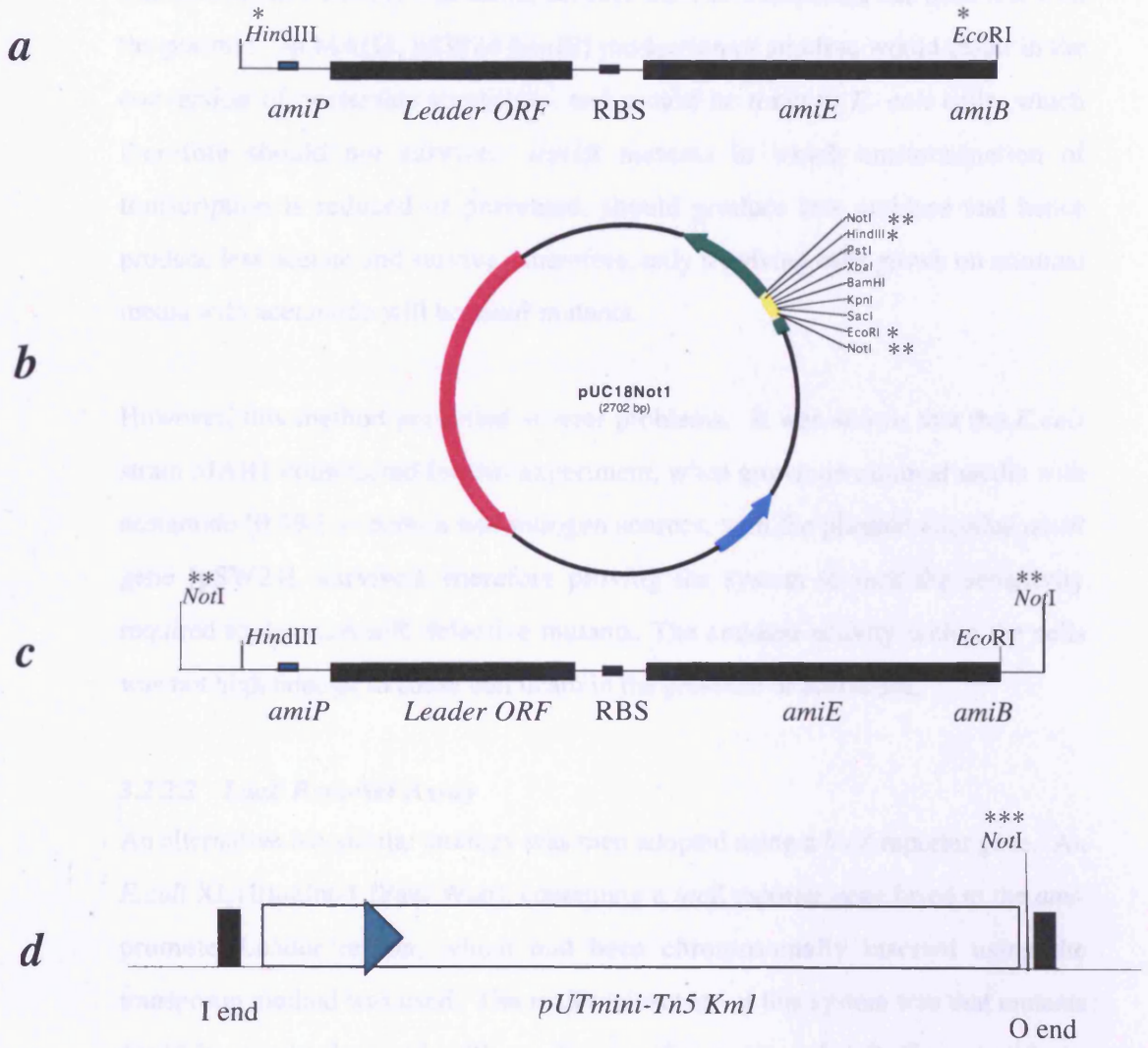


Figure 3.3. Stages in the Construction of strain MAH1

[a] The amiE/Leader gene fragment was liberated from pMH1 on a HindIII/EcoRI fragment [*] and ligated into pUC18NotI to create pMH10 [b]. [c] The NotI [**], amiE/Leader fragment was subsequently excised from pMH10 and cloned into [d] pUTminiTn5 Km1 to create the donor plasmid pMH11.

The new strain – MAH1 was stable because the Tn5 transposase has been lost with the plasmid. In MAH1, pSW24 [*amiR*] production of amidase would result in the conversion of acetamide to acetate, and would be toxic to *E. coli* cells, which therefore should not survive. *amiR* mutants in which antitermination of transcription is reduced or prevented, should produce less amidase and hence produce less acetate and survive - therefore, only surviving cells grown on minimal media with acetamide will be *amiR* mutants.

However, this method presented several problems. It was shown that the *E.coli* strain MAH1 constructed for this experiment, when grown on minimal media with acetamide [0.3%] as carbon and nitrogen sources, with the plasmid encoded *amiR* gene [pSW24], survived, therefore proving the system to lack the sensitivity required to detect AmiR defective mutants. The amidase activity within the cells was not high enough to cause cell death in the presence of acetamide.

3.2.2.2 LacZ Reporter Assay

An alternative but similar strategy was then adopted using a *lacZ* reporter gene. An *E.coli* XL1BlueInt-1 [Paul Wan], containing a *lacZ* reporter gene fused to the *ami*-promoter/Leader region, which had been chromosomally inserted using the transposon method was used. The major advantage of this system was that mutants could be easily detected with an X-gal colour screen for β–Galactosidase. However, this system presented its own problems. This screen also lacked the sensitivity for detecting AmiR defective mutants, as the assays with the WT AmiR again produced low levels of β–galactosidase activity. The systems discussed here thus far, both utilised a method whereby only a single copy of *amiE/Leader* sequence exists in the cell.

3.2.2.3 A Two Plasmid System

A third method was tried: A system whereby *amiR* was expressed from a high copy number plasmid and with *amiE* [pTM1] on a second plasmid was employed. As shown previously [Table 3.2], the strain with only pTM1 give an amidase SA of 1.0 and the presence of amiR [pSW24] leads to constitutive antiterminated amidase activity of 37.0.

3.2.2.4 Hydroxylamine Mutagenesis of AmiR

For the mutagenesis of *amiR*, hydroxylamine [NH_2OH] was used [Section 2.2.2.15]. Hydroxylamine is a well-known mutagenic agent which reacts with the amino groups of the nucleic acid bases cytosine and adenine, giving rise to promutagenic N4-hydroxycytosine and N6-hydroxyadenine. Both N-hydroxy compounds can base pair not only like the parent cytosine and adenine but also like uracil and guanine. Hydroxylamine mutagenesis thus results in C to T and G to A transition mutations when double stranded DNA is mutagenized. This method was utilised to produce pSW24 mutants – and *amiR* defective mutants were identified by carrying out amidase assays in *E.coli* cells bearing *amiE* on a second plasmid [pTM1] – those with reduced amidase SA less that the WT *amiR* of 37.0 were sent for sequencing to identify the sequence change.

This strategy proved to be very unreliable which was possibly due to the fact that two plasmids were being used. Strains were isolated in which amidase activity was much reduced, however isolation and sequence determination of the pSW24 plasmids from these strains showed no difference from WT pSW24.

3.2.3 *Development of Dual Expression of amiR and amiE from a Single Plasmid*

To retain the sensitivity, but without the need for a two plasmid system, a method whereby two proteins could be expressed from a single high copy number plasmid was adopted. The pETDuet-1 cloning vector is specifically designed for the co-expression of two genes under the control bacteriophage T7 transcription and expression of which is induced by providing a source of T7 polymerase in the host cell [Novy *et al.*, 2002; Dubendorff, and Studier, 1991]. The pETDuet-1 vector contains two MCS's, both of which are preceded by a T7 promoter /*lac* operator and RBS [Figure 3.4]. Co-expression of multiple target genes in *E. coli* is advantageous for studying protein complexes and protein–protein interactions and this pET System was used to detect *amiR* mutations by carrying out amidase assays as a measure of the antitermination activity by the mutated AmiR protein. This system provided a simple solution to the problem of transforming two separate plasmids into the host cell.

3.2.3.1 Creation of the AmiR and AmiE Dual Expression Vectors

The *amiE*/leader sequence was subcloned from pMH1 [Section 5.1.2.1.] on a 2.5kb *Hind*III/*Eco*RI fragment and ligated into pETDuet1 MCS1 giving pMH90 [Figure 3.5]. The orientation of this fragment in MCS1 was such that the *amiE* leader sequence was not under the control of the T7 promoter upstream of MCS1. This was favourable as high levels of expression from the T7 promoter combined with the physiological requirements for the antitermination of the leader sequence, which allow for low levels of read through in the absence of AmiR, could otherwise generate high levels of background amidase activity and subsequent false positive results.

Expression of *amiE* in pMH90 and pMH901 is dependent on transcription by the host strain's native RNA polymerase initiating from the *ami* promoter. AmiR was

PCR amplified from pSW24 with primers containing *Nde*I [forward] and *Xho*I targets [reverse primer] [Table 3.3.]. The subsequent PCR product [Figure 3.6.] was digested with *Nde*I and *Xho*I restriction enzymes and ligated into the MCS2 of pMH90, thus creating pMH901. Simultaneously the *amiR* PCR product was cloned into the MCS2 of pETDuet-1 giving pMH91. pMH901 is the plasmid in which this system can be tested with pMH90 and pMH91 providing controls.

Ten of the potential pMH901 and pMH91 colonies were picked following ligation and transformation and the insertions were confirmed by sequencing using the pETUpstream Primer for *amiE* sequence cloned into MCS1 and the DuetUP2 Primer to confirm presence of the *amiR* sequence cloned into MCS2 [Table 3.4.].

Table 3.3. Primers used for PCR of AmiR in this study.

Name	***Sequence [5' –3']***	***Features***
AmiRFor1	**GAGGCATATG**AGCGCCAACTCGCTG	*Nde*I [CATATG] **Initiator ATG**
AmiRRev1	TTAA**CTCGAG**TCAGGCGGACGGCTC	***Xho*I [CTCGAG]**

Table 3.4. Sequencing primers for the pETDuet Vectors [Novagen].

Name	***Sequence [5' –3']***
pETUpstream Primer	ATGCGTCCGGCGTAGA
DuetUP2 Primer	TTGTACACGGCCGCATAATC

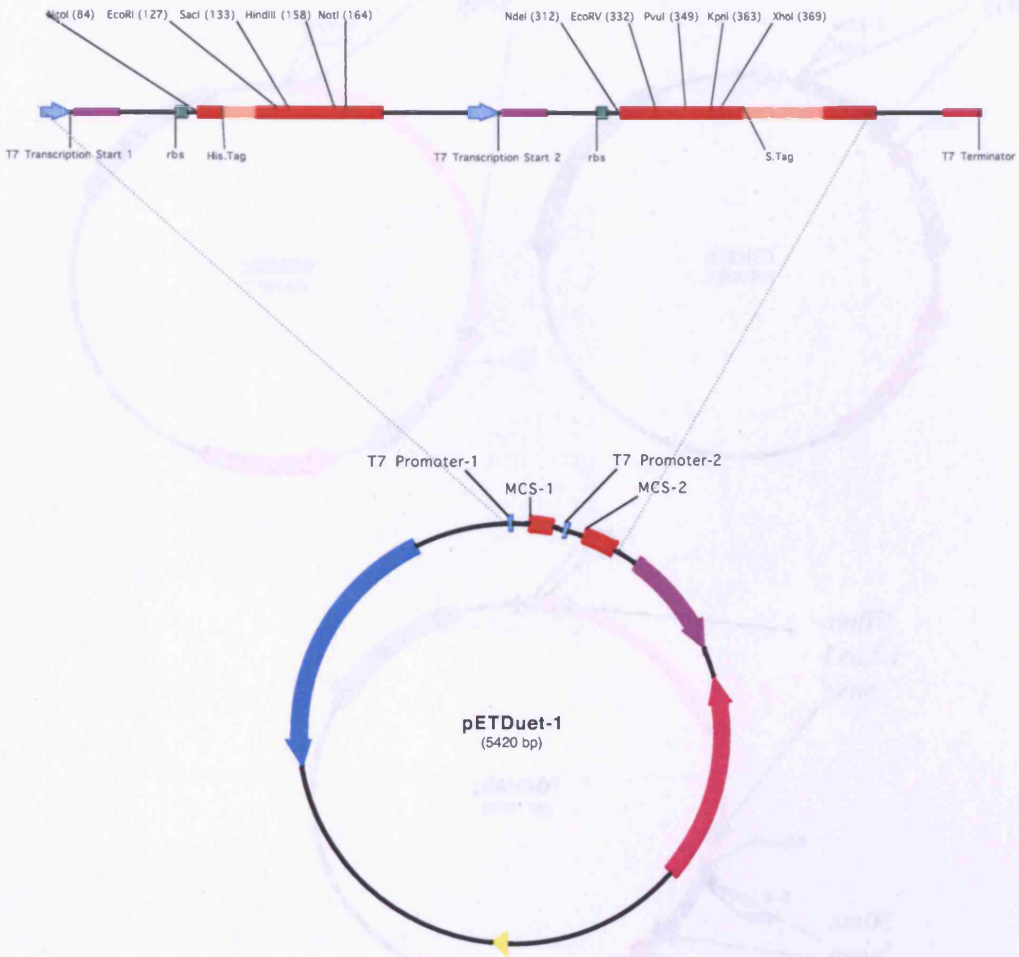


Figure 3.4. pETDuet-1 plasmid map

Plasmid map of pETDuet-1 showing the multiple cloning sites 1 and 2. Above this, the MCS regions showing some of the target sites used for cloning. The vector encodes two multiple cloning sites [MCS] each of which is preceded by a T7 promoter/lac operator and ribosome binding site [rbs]. The vector also carries the pBR322-derived ColE1 replication origin [yellow], lacI gene [aqua-blue] and ampicillin resistance gene [magenta].

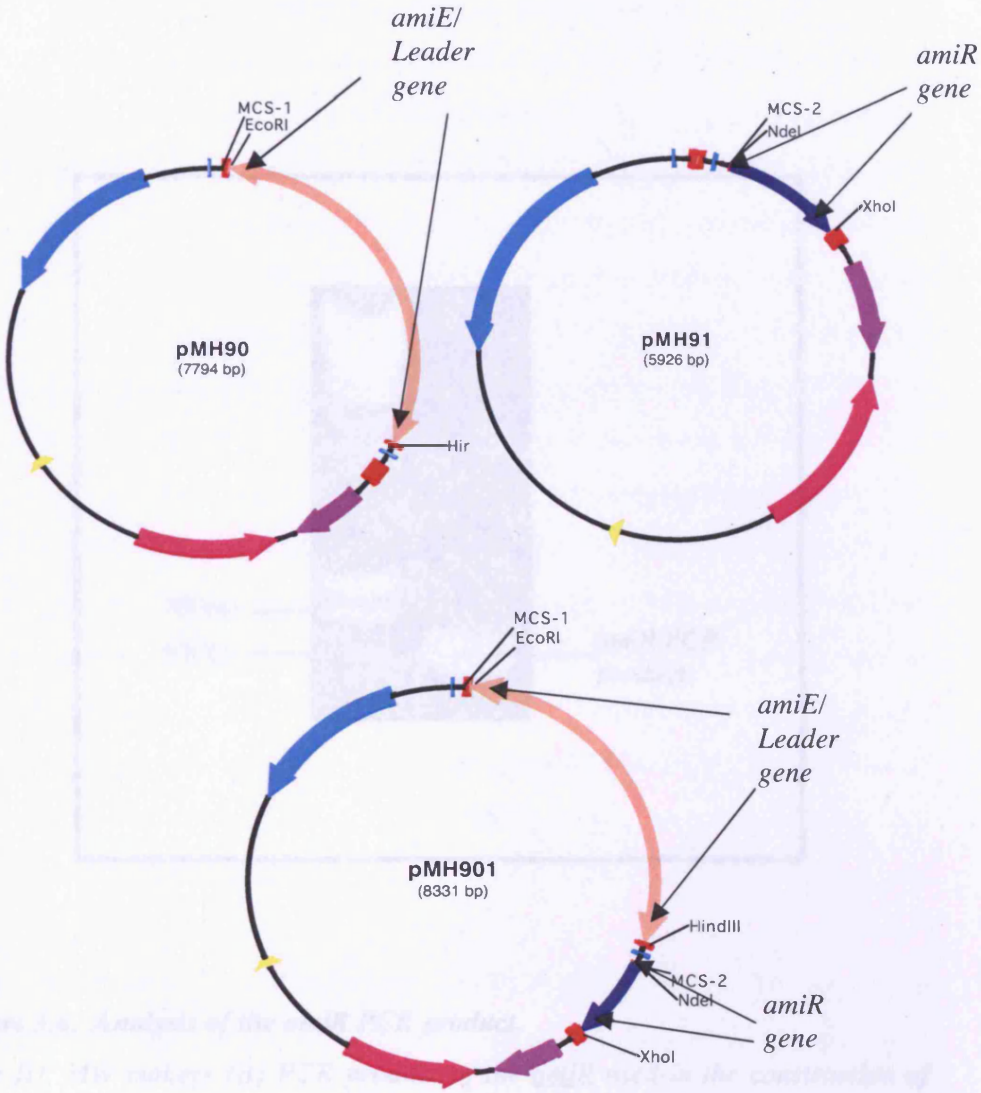


Figure 3.5. Structural maps of pMH90, pMH91 and pMH901.

pMH90 and pMH901, have the amiE gene and leader sequence cloned into the HindIII-EcoRI site of MCS1. The fragment was cloned into the MCS1 in the opposite direction of T7 transcription to allow for independent transcription of the amiE gene to enable the monitoring of AmiR activity by transcription antitermination. pMH91 and pMH901 have the WT amiR gene sequence cloned into the MCS2.

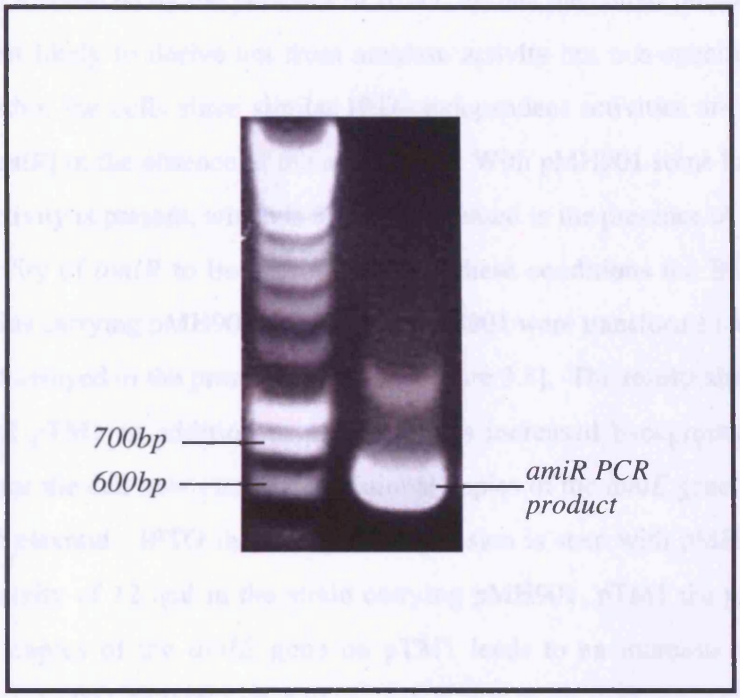


Figure 3.6. Analysis of the amiR PCR product.

Lane [i], MW makers [ii] PCR product of the amiR used in the construction of pMH90/pMH901.

3.2.3.2 Amidase Assays with the Dual Expression Vector

The results of the amidase assays using pMH90, pMH91 and pMH901 are shown in Figure 3.7. The strain containing pMH90 shows background levels of amidase activity not influenced by the presence of IPTG. In fact the colour production seen here is most likely to derive not from amidase activity but non-specific esterase activity within the cells since similar IPTG-independent activities are seen with pMH91 [*amiR*] in the absence of the *amiE* gene. With pMH901 some background amidase activity is present, which is 3.2 fold increased in the presence of IPTG. To test the ability of *amiR* to be expressed under these conditions the BL21 [DE3] pLysS strains carrying pMH90, pMH91 and pMH901 were transformed with pTM1 [*amiE*] and assayed in the presence of IPTG [Figure 3.8]. The results show that the presence of pTM1 in addition to pMH90 gives increased background amidase activity since the cell now contains additional copies of the *amiE* gene present on the second plasmid. IPTG induced gene expression is seen with pMH91, pTM1 with an activity of 12 and in the strain carrying pMH901, pTM1 the presence of additional copies of the *amiE* gene on pTM1 leads to an increase in amidase activity from 13.0 to 18.0 [Figure 3.8].

Thus, *amiR* expression is IPTG inducible in BL21 [DE3] pLysS. Plasmid pLysS constitutively expresses low levels of T7 lysozyme, which reduces the basal expression of recombinant AmiR by inhibiting basal levels of T7 RNA polymerase. In the absence of IPTG in these cells, some leaky *amiR* expression occurs, but is tightly regulated. The T7 RNA polymerase of bacteriophage T7 has a stringent specificity for its own promoters, which do not occur naturally in *E. coli*. Studier and Moffat [1986] showed that relatively small amounts of T7 RNA polymerase provided from a cloned copy of T7 gene 1 is sufficient to direct high-level transcription from a T7 promoter in a multicopy plasmid. Such transcription can proceed several times around the plasmid without terminating, and can be so active that transcription by *E. coli* RNA polymerase is greatly decreased. Strain

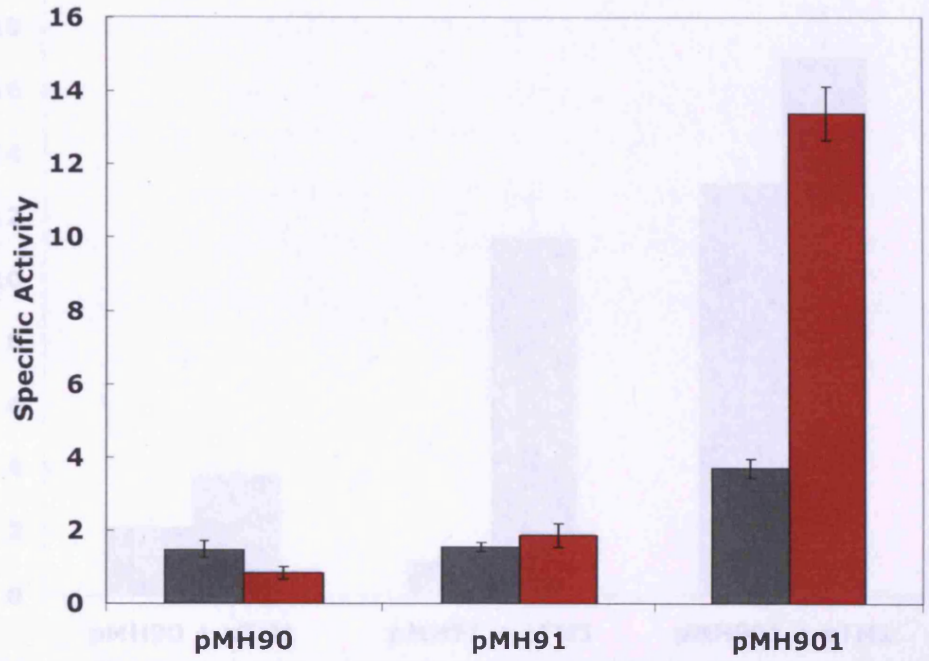


Figure 3.7. Amidase assays of E.coli BL21 DE3 pLysS containing the pETDuet-1 derived constructs.

The chart shows the specific activity values of the cells in the absence of IPTG [grey] and presence of IPTG [red]: pMH90 [1.5 (-IPTG) and 0.8 (+IPTG)], pMH91[1.5 (-IPTG) and 1.9 (+IPTG)] and pMH901 [3.7(-IPTG) and 13.4 (+IPTG)].

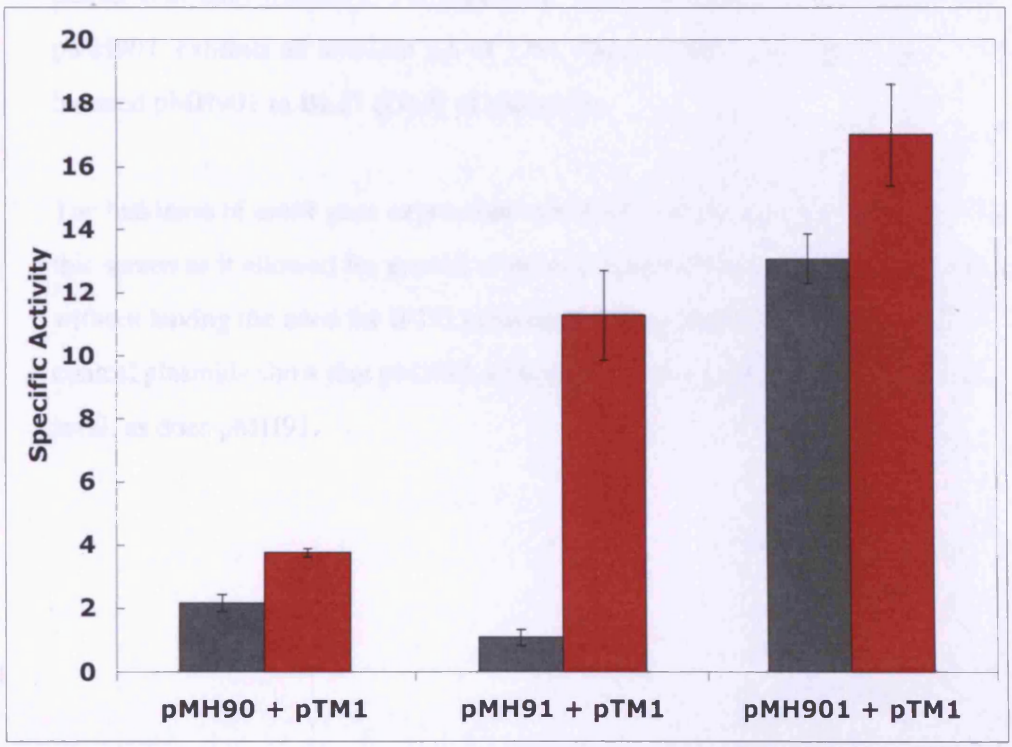


Figure 3.8. Amidase assays of E.coli BL21 DE3 pLysS containing IPTG induced pETDuet-1 derived constructs.

The chart shows the SA values of induced cells in absence [grey] and the presence [red] of amiE [pTM1] to test the activity of the amiR gene cloned into pMH91: pMH90 [SA 2.2 (- amiE) and SA 3.8 (+ amiE)], pMH91 [SA 1.1 (- amiE) and SA 11.3 (+ amiE)], and pMH901 [13.1 (- amiE) and 17.0 (+ amiE)].

BL21 [DE3], carries the IPTG inducible T7 RNA polymerase gene on a lysogenic prophage.

Amidase assays were also carried out with *E.coli* BL21 [DE3], pMH90, where pLysS is absent, without IPTG induction. Amidase assay results of BL21 [DE3], pMH901, exhibits an amidase SA of 13.0, which is almost as high as in the IPTG induced pMH901 in BL21 [DE3] pLysS strain.

The leakiness of *amiR* gene expression exhibited in BL21 [DE3] was beneficial to this screen as it allowed for growth of several hundreds of clones in 96-well plates, without having the need for IPTG induction. Specific activity results of the control plasmids show that pMH90, as expected, has a very low specific activity level, as does pMH91.

3.2.4 Random mutagenesis of amiR using Error prone PCR

Error prone PCR is another method which is frequently used for generating amino acid substitutions in proteins by introducing random mutations into a gene during error prone PCR reactions. Incorporation of an incorrect base pair during the DNA amplification in PCR reactions are generally infrequent and can be avoided if a proof reading DNA polymerase is used [Bedinger and Alberts, 1983]. Non-proof reading error prone DNA polymerases can be used to increase the likelihood of PCR born mistakes. The products of such PCR reactions can then be cloned into an expression vector and the resulting library screened for novel structural and functional changes in the protein.

Taq DNA polymerase is a non-proof reading polymerase and can be forced into introducing errors into DNA if it is buffered by Mn^{2+} ions and unbalanced levels of nucleotides are present [Shafikhani *et al.*, 1997]. This method however creates a mutational bias which can hinder a functional screen of a protein where it is beneficial to screen all possible amino acid changes. For this study the commercially available Genemorph II Random mutagenesis kit [Stratagene] was used for the amplification of a library of mutagenic *amiR* clones. This kit was chosen because it uses a commercially developed DNA-pol, Mutazyme II which exhibits less mutational bias than the procedure described by Shafikani [1997].

Like most error prone DNA polymerases, Mutazyme II, favours transition mutations over transversions but it introduces transition mutations [AT>GC and GC>AT] at a similar rate [AT>GC/GC>AT ratio =0.6] unlike *Taq* which is twice as likely to introduce AT>GC transition over GC>AT. Furthermore Mutazyme II introduces mutations at A's and T's only slightly more frequently than G's and C's whereas *Taq* is four times more likely to do this. It has also been shown that altering the buffering conditions and dNTP ratios of the PCR reactions, a requirement for altering the rate of mutation using *Taq*, can further enhance these

mutational biases. The GeneMorph II kit uses balanced dNTP concentrations and constant $MgCl_2$ buffer conditions, generating varying mutation rates with the alteration of starting template concentrations and PCR cycle numbers. Increasing template DNA concentration and lowering the cycle number reduces the mutation rate of the reaction, high mutation frequencies are achieved with low template concentrations and high cycle number. The spectrum of mutations remains constant at all mutation frequencies with the Genemorph II kit.

3.2.5 Creating a library of mutated amiR genes within a pMH901 plasmid.

The advantages of having both *amiE*/Leader and *amiR* on the same plasmid for performing amidase assays have been shown in section 3.2.3.2. This was developed so that once the desired mutation frequency of the *amiR* gene was achieved by PCR the resultant PCR products would be directly ligated into this one plasmid pETDuet-1 derived system. When randomly mutagenising a protein to determine novel areas of interest in relation to both structure and function, informed conclusions can only be made from single amino acid changes. In this screen the level of mutation was hence adjusted so that only single mutations were introduced into the *amiR* gene clones. As the *amiR* gene is only 569 bp, the conditions of the mutagenic PCR were for the introduction of 2 mutations/kb of DNA. The recommended low range mutation rate of 0-4.5 mutations/kb of the kit were initially followed [100ng of template DNA and 20 cycles of the PCR]. The *amiR* template DNA was PCR amplified from pMH91, using *AmiRFor1* and *AmiRRev1* [Table 3.3] primers and Expand™ DNA polymerase. The resultant product was run on an agarose gel to check it was the correct size before quantifying the DNA concentration spectrophometrically at 260nm. 100ng of this template was then amplified with the same primers using Mutazyme II polymerase [Figure 3.9.], digested and ligated into the *Nde*1 *Xho*1 cleaved pMH90. The product of this

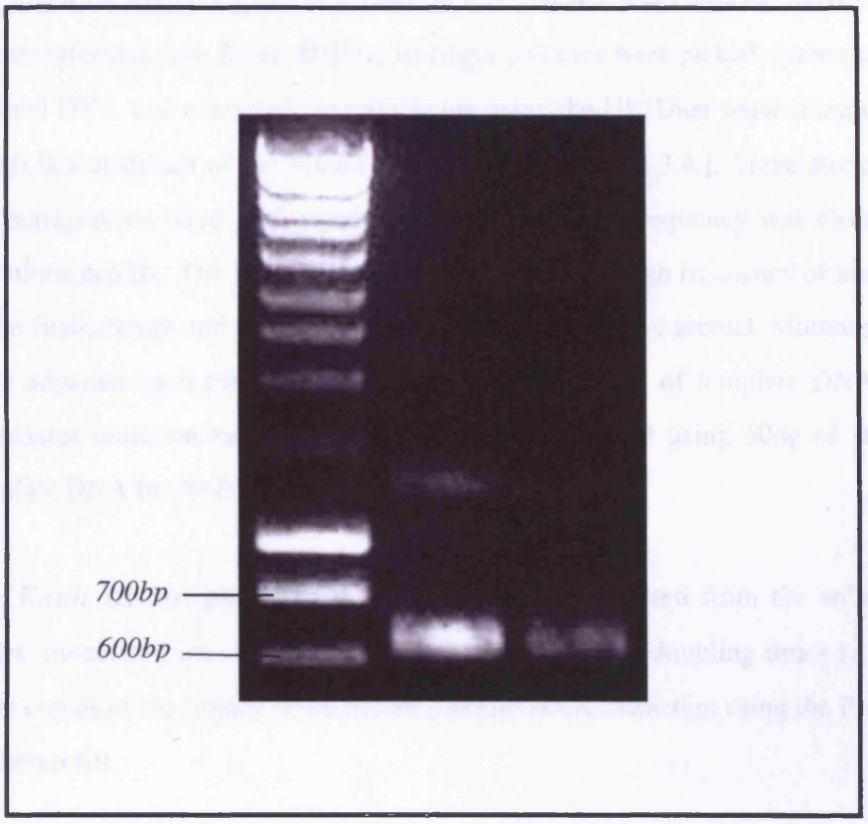


Figure 3.9. Agarose gel electrophoresis of the WT and mutagenised amiR PCR products.

Lane [i], MW makers [ii] PCR product of the amiR used in the construction of pMH90/pMH901. [iii] the mutagenic amiR PCR product.

ligation was a library of mutated *amiR* held on a plasmid containing *amiE*/Leader and was called pMH901[M].

The mutation frequency of *amiR* held on this plasmid was checked, pMH901[M] was transformed into *E.coli* DH5α, 10 single colonies were picked, grown up and plasmid DNA was extracted for sequencing using the UP2Duet sequencing primer which lies upstream of the MCS2 of pMH901[M] [Table 3.4.]. Three attempts at the mutagenesis were performed before the mutation frequency was close to 2 mutations per/kb. The intial two attempts failed with a high frequency of mutation in the first attempt and too low a mutation frequency in the second. Mutation rates were adjusted each time by lowering the concentration of template DNA. The successful mutation rate of pMH901[M3] was achieved using 50ng of starting template DNA for 20 PCR cycles.

The *E.coli* DH5α, pMH901[M3] colonies were harvested from the solid agar plates, inoculated into liquid culture and grown for four doubling times to create more copies of the library DNA before plasmid DNA extraction using the Promega Midiprep Kit.

To perform the amidase assays and assess the mutations of *amiR* pMH901[M3] DNA was transformed into competent BL21 DE3 cells.

3.2.6 A Large Scale Amidase Assay Screen of the pMH901[M3] library.

To isolate clones containing important amino acid changes from the library of mutated *amiR* genes now held on pMH901[M3], strains of Bl21 [DE3] pMH901[M3] were screened using amidase activity.

The amidase assay was therefore modified so it could be performed in single wells of a flat bottomed 96-Well microtitre plate i.e. in volume less than 500 µl. Figure 3.9 [a], shows a time course of amidase transferase activity in a wells of a 96-well plate, containing BL21 [DE3] strain transformed with WT pMH901. Expression of *amiE* is permitted by the expression of functional *amiR*. At 5 minutes the colour change from yellow to dark red was clearly visible. Beyond this time point the colour change is less detectable from dark red to black. A screen of the mutant library was subsequently set up in multiple 96-well plates whereby the activity of mutated AmiR clones were to be tested.

Several hundred colonies from the transformed BL21 [DE3] cells were patched out onto Amp plates. For the assay, cells from each patch were inoculated into 300 µl of LB + Amp in 9 flat bottom 96-well plates and grown overnight at 37 C, without shaking. Controls were grown in the lane 12, as shown in figure 3.10 [b]. 10 µl of overnight culture was then dispensed into 90 µl of mixed substrates at 15 second intervals using a multi channel pipette and incubated for five minutes at room temperature. Reactions were stopped by addition of 200 µl of ferric chloride-HCL solution and the results read by eye. The results are shown in Figure 3.11 and samples which showed reduced amidase activity were analysed further. Of the 792 colonies analysed using this system, 117 were deemed to have reduced activity [when compared the WT AmiR produced in wells 12 A-D of the corresponding plate] and 81 were subject to further analysis by conventional amidase assays [Table 3.5].

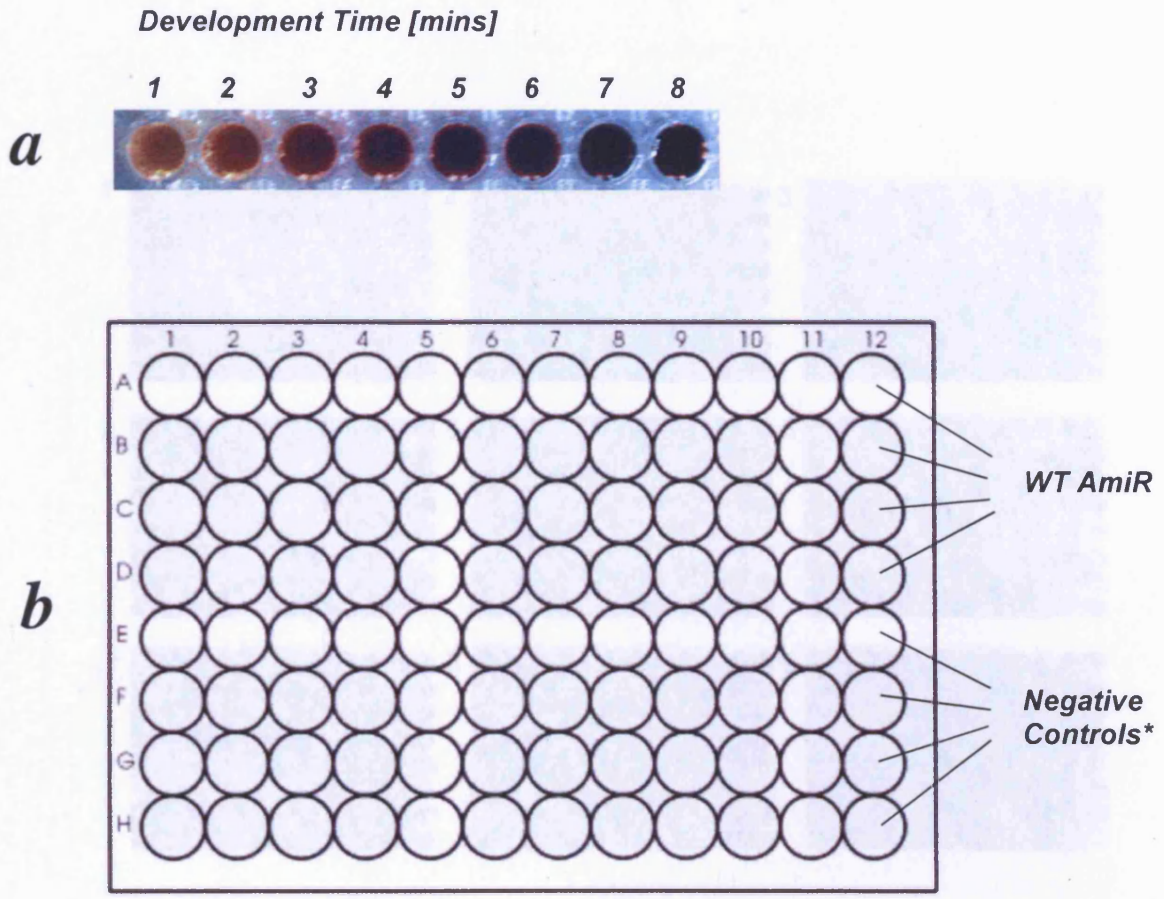


Figure 3.10. Organisation of the 96-Well Plate Mutant Screen.

[a] Time course of WT AmiR amidase assays of 300 µl volume in microtitre wells. [b] Layout and numbering of the 96-well plate. In every instance, lane 12 A-D contained pMH901 [WT amiR + amiE/ Leader/], 12 E-F contained pMH91 [WT amiR only], 12 G contained assayed pMH90 [ami/Leader/amiE only], and 12 H contained a reagent blank. Each plate allows 88 potential mutants to be screened prior to being selected for a full amidase assay.

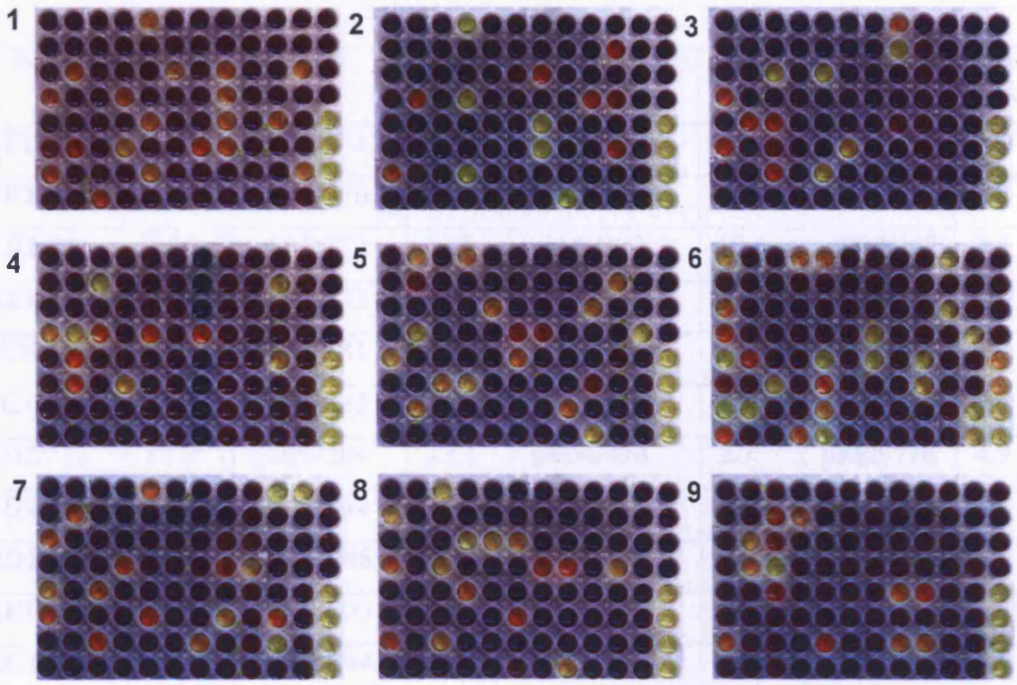


Figure 3.11. The 96-Well Plate Mutant Screen.

The nine plates used in the screen [numbered top left hand corner]. From these plates, colonies producing visibly less amidase activity compared to the WT amiR [wells 12 A-D], were selected for further tests.

Table 3.5. Second phase of the mutant selection.

Strains which showed reduced activity in the 96-well screen were tested using the conventional amidase assay. Specific activity levels are the average results of amidase assays carried out in duplicate. Strains were named to aid identification; p followed by plate number and location / corresponding patch from the original colony. Blue highlighted strains were selected for sequence analysis.

Strain	Specific Activity	Strain	Specific Activity	Strain	Specific Activity	Strain	Specific Activity
p1F1/56	**8.9**	p3G9/251	13.0	p5B2/365	10.7	**p7E9/681**	**9.3**
p1F2/57	**6.4**	p4A6/270	12.7	**p5D6/391**	**6.4**	**p7F5/588**	**3.0**
p1H1/88	**2.8**	p4A8/272	13.8	p5D7/392	10.4	**p7G3/597**	**3.6**
p2B10/109	**8.6**	p4A9/273	12.4	**p5D10/395**	**9.2**	p8D7/656	10.0
p2C6/116	**8.5**	p4B6/281	10.6	p5E10/406	12.1	**p8D8/657**	**6.7**
p2C7/117	13.8	p4B8/283	12.8	**p5F9/416**	**7.1**	**p8F6/677**	**9.5**
p2D2/123	11.0	p4B9/284	13.1	**p6D1/474**	**2.7**	**p9B3/718**	**4.9**
p2D9/130	11.7	p4C8/294	11.8	p6E2/486	10.5	**p9C3/729**	**7.2**
p2D10/131	**7.3**	**p4C9/295**	**9.7**	**p6E4/484**	**9.8**	**p9E3/751**	**8.1**
p2F10/153	**5.7**	p4D3/300	11.7	**p6F1/496**	**7.3**	p9E9/757	11.7
p2G1/155	**4.0**	**p4D7/304**	**5.9**	p6F2/497	12.9	**p9F3/762**	**9.7**
p3B9/196	**5.6**	p4D8/305	10.7	p6F4/499	13.0	p9G2/772	12.9
p3C9/207	**7.7**	p4D9/306	13.1	**p6F5/500**	**2.4**	p9G3/773	12.1
p3D1/210	**3.3**	p4E2/310	1.7	p6G3/509	11.6	**p9G5/775**	**6.0**
p3E2/222	12.1	p4E8/316	11.9	p6G4/510	12.9	**p9H3/784**	**8.8**
p3E3/223	10.3	p4E9/317	13.0	p6H2/519	12.4	**pMH90**	**1.7**
p3E8/228	10.8	p4F8/327	12.5	p6H3/520	13.3	**pMH91**	**1.1**
p3E9/229	13.7	p4F9/328	13.6	p7A5/533	1.8	**pMH901**	**12.9**
p3F1/232	**4.2**	**p4G8/338**	**9.8**	**p7C8/558**	**7.3**		
p3F3/234	10.9	p4G9/339	12.7	**p7C9/559**	**9.8**		
p3F9/240	12.8	p4H8/349	12.9	p7D4/565	12.3		
p3G3/245	12.6	p4H9/350	13.3	p7D8/569	11.4		

3.2.7 Further Analysis of AmiR Clones Exhibiting Reduced Amidase Antitermination Activity

The amidase activity profiles from the proposed *amiR* mutants varied from 1.7 [background] to 13.8 [full activity]. The selection criterion for further analysis was if the specific activity levels were significantly lower than WT AmiR, which was 13.2. Those selected had activity levels of less than 10.0 and greater than 1.5, [which is the usual background level produced by *amiE* (pMH90) alone].

3.2.7.1 Sequencing of Selected Mutants

DNA from 35 of the potential mutant clones was sent for sequencing. Of these, only three of the recombinants had one mutation within the *amiR* gene. Several had no mutations at all and the remainder had conservative mutations. An example of this was the mutation in p7C9/559 of G to T in the coding of Leucine16: CTG > CTT, which both code for Leucine.

3.2.7.2 AmiR Mutants Isolated in this Screen were Found in the NTD and CT-ANTAR Domain of the Protein.

The three clones isolated in this screen which result in a single amino acid change in AmiR were; p3F1/232, p6D1/474 and p6F5/500 which gave specific activity values of 4.2, 2.7 and 2.4 respectively. Sequencing of p3F1/232 revealed a mutation which resulted in the amino acid change of Proline 20 to Glutamine [CCG>CAG] [herein referred to as pAmiR901P20Q]. The mutation found within *amiR* in p6D1/474 was located at position 154 resulting in the amino acid change Valine 154 to Glutamate [GTG>GAG], [herein referred to as pAmiR901V154E]. Sequencing from p6F5/500 revealed a mutation resulting the amino acid change of Arginine 179 to Histidine [CGC>CAC] [herein referred to as pAmiR901R179H].

3.2.8 Summary of Chapter 3

The first aim of this chapter was to test the effects of mutating the conserved residues found within the putative RNA-binding domain of AmiR [ANTAR] and secondly to isolate and identify key residues within AmiR which are important for the transcription antitermination reaction *in vivo* by using a system whereby antitermination defective AmiR mutants are selected for in a random mutagenesis screen.

To achieve the first goal, SDM was used. When carrying out SDM to a protein, it is necessary to consider the following; [i] which amino acids in the protein are of interest? and [ii] which residue changes should be made? The first question in the case of the AmiR protein was simple, and the residues which has been shown to be strictly conserved in the proposed RNA-binding domain [ANTAR] were selected [Figure 3.2], [Shu and Zhulin, 2002]. Residues Alanine 152, Alanine 167 and Histidine 168 were successfully targeted for SDM. Attempts were made to alter residues Alanine 175 and Alanine 186 however, after several failed tries to achieve the desired mutation, efforts were abandoned due to time constraints.

Having selected these residues, it was then necessary to decide on the amino acid to replace them. In general, the replacement should be to a residue which is of a similar size, but with a different chemical property i.e. charge [Plapp, 1995]. The selected Alanine residues were changed to Glycine and/or Valine and the Histidine residue changed to Alanine, positively charged R and negatively charged E.

To achieve the second goal of isolating antitermination defective *amiR* mutants, random mutagenesis was used which involved using an error prone DNA polymerase in the PCR reaction to incorporate mutations throughout the *amiR* gene. Before this library of mutant proteins could be screened, a strategy in which

the mutated AmiR proteins could be expressed from a single plasmid which also carries the *ami*/leader had to be developed and tested. Mutagenised AmiR constructs were then assayed for amidase activity, and the results summarised in Table 3.6.

Interpretation of mutation studies is often difficult because amino acid replacements can affect folding pathways and subunit interactions as well as ligand interactions. Of the changes made here it is most apparent that residue Histidine 168 is important for the RNA interaction during antitermination since replacement by an alternative positively charged residue [R] or negatively charged residue [E] caused complete loss of antitermination activity. The results of the changes to A152 and A167 are more difficult to interpret since it is possible that they will lead to structural changes and some loss of activity because of this.

Table 3.6. Summary of the AmiR mutants generated in this study

Construct	*Amino Acid Change*	*Specific Activity*	*WT Specific Activity*
pAmiR152A>G	A > G	18.6	37.0
pAmiR152A>V	A > V	22.0	37.0
pAmiR167A>G	A > G	2.4	37.0
pAmiR168H>A	H > A	11.35	37.0
pAmiR168H>R	H > R	1.0	37.0
pAmiR168H>E	H > E	0.9	37.0
pAmiR901P20>Q	P > Q	4.2	13.0
pAmiR901V154>E	V > E	2.7	13.0
pAmiR901R179>H	R > H	2.4	13.0

The results from the random mutagenesis experiments uncovered three residues where changes to which resulted in much reduced activity. In the system used here, WT AmiR levels were 13.0. Mutations V154E, and R179H both resulted in an approximate 5 fold reduction in amidase activity. Mutation P20Q, resulted in a 3 fold reduction in amidase activity. The V154E mutation affects a residue buried between the two AmiR molecules towards the bottom of the long coiled coil, within ANTAR and makes major contact with M157 and E164 on the other dimer member. The change of residue from hydrophobic to hydrophilic within this region might be expected to lead to instability of the dimer [Figure 3.12]. The R179H mutation lies within a group of charged residues [KRRE] making the turn from helix 7 to helix 8. The P20 residue in the NTD of AmiR,is found with the sequence LNPPG in the loop between the end of a β-strand a and beginning of helix 2 [Figure 3.13]. The P20Q change identified may cause a structural change to the NTD.

As the system utilised here does not involve induction or is not subject to negative protein-RNA interaction, or exist to support the structure of the active protein. The data presented here provided further support to the theory that ANTAR has RNA binding activity through the discovery of two novel residues within ANTAR which alter antiterminatory effects of AmiR when mutated.

However, this method could be improved upon by increasing the accuracy of the amidase assays results when selecting mutants for further analysis in the second phase of the screen [assays were carried out in duplicate only – and those which had wild type *amiR* sequences should not have been detected in this screen]. Alternatively, only those strains which had much reduced amidase activities [SA's of 1.0 – 5.0] should have been selected, although the possibility exists that many of.

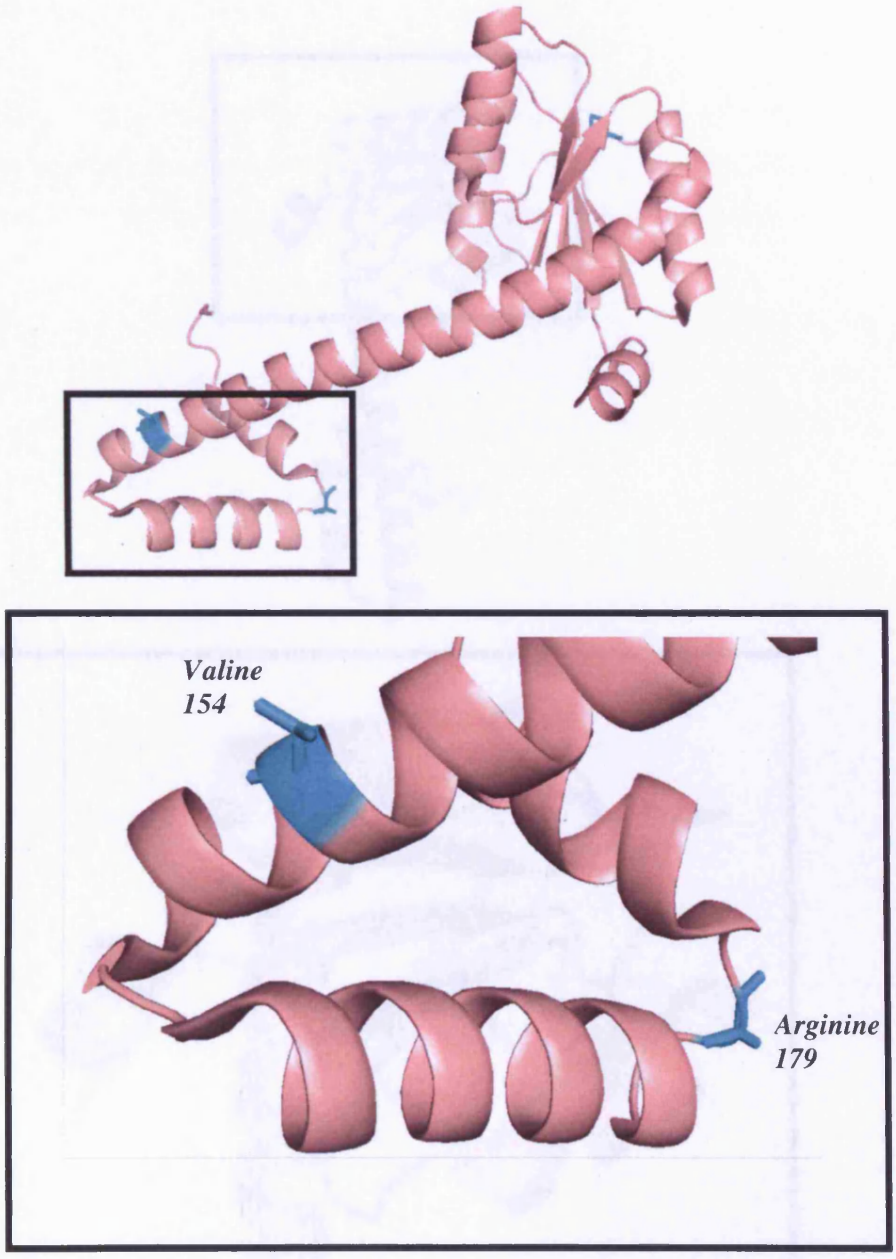


Figure 3.12. Mapping the locations of V154 and R179 residing within the ANTAR domain of AmiR.

Each of the randomly generated V154E and R179H mutations uncovered in this screen exhibit a 5 fold reduction in amidase activity compared to WT AmiR.

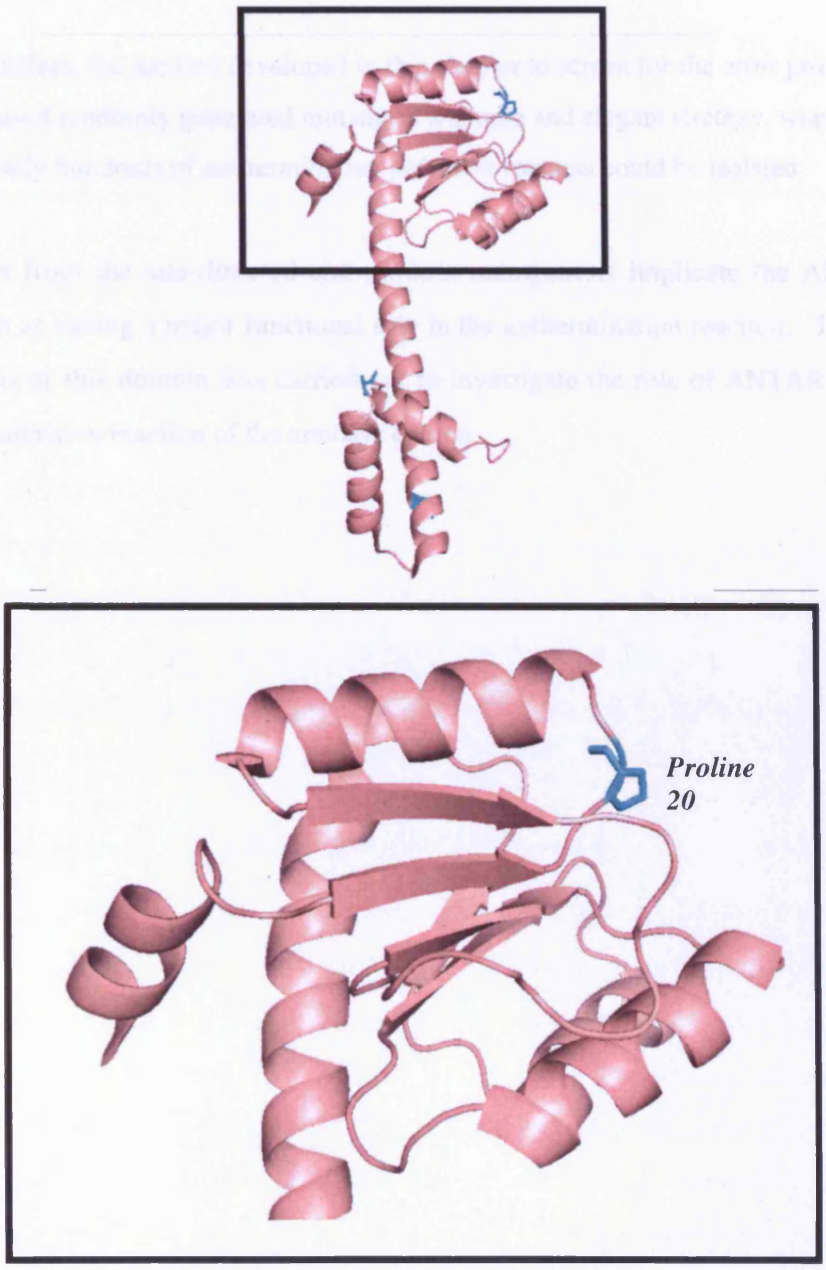


Figure 3.13. Mapping the location of P20 located within the NTD of AmiR.

The randomly generated P20Q amiR gene mutation isolated in this screen, resulted in a 3 fold reduction in amidase acivity.

those strains would have more than one mutation, making interpretation of the result difficult

Nevertheless, the method developed in this chapter to screen for the error prone PCR based randomly generated mutants is a simple and elegant strategy, whereby potentially hundreds of antitermination defective mutants could be isolated.

Results from the site-directed and random mutagenesis implicate the ANTAR domain as having a major functional role in the antitermination reaction. Further analysis of this domain was carried out to investigate the role of ANTAR in the antitermination reaction of the amidase operon.

4 *Purification of ANTAR, the Putative RNA-binding domain of AmiR*

4.1 *Expression and purification of the ANTAR Domain*

4.1.1 Introduction

In the crystal structure of the AmiC/AmiR- butyramide complex AmiR exists as a dimer interacting with two AmiC monomers. The C-terminus of AmiR where the ANTAR domain is located folds as a three-helix bundle. Mutation analysis carried out previously, and in this study, implicate the ANTAR domains involvement in the antitermination reaction and it is highly likely to have RNA binding activity.

4.1.2 Rationale for cloning the ANTAR Domain

Inducing conditions, such as the binding of acetamide to AmiC, weakens the AmiC AmiR interaction such that the affinity of AmiR for the leader RNA will be higher than that for AmiC. It is presently unclear how AmiC exerts its negative regulatory effect other than the fact that it sterically hinders AmiR by complex formation and prevents it from binding leader RNA. It is possible that the role of AmiC is to either prevent oligomerisation-induced activation of AmiR, or to prevent an internal AmiR dimer rearrangement. Thus, should the ANTAR domain be expressed as its own entity, this would be representative of the domain in the form that is amenable to RNA-binding. Expression of the ANTAR domain on its own would allow investigation of possible oligomerisation states of the molecule. It would then be

possible to investigate the mode of RNA binding, its affinity and to investigate if the binding is sequence specific.

4.1.2.1 Hydropathicity Plots of AmiR Indicate ANTAR is Highly Soluble

Previous investigations of AmiR had been hampered by the fact that expressed by itself in *P. aeruginosa* or *E.coli*, although able to cause antitermination, the protein aggregates upon overproduction. It is only soluble, in a polymeric form, under conditions of high salt and high pH. Hydropathicity analysis of AmiR and ANTAR show that the ANTAR domain in mainly hydrophilic and therefore should be soluble [Figures 4.1. and 4.2.]. Hydropathy plots were generated using the Kyte and Doolittle scale [Kyte and Doolittle, 1982]. This scale is used to highlight hydrophobic regions within a protein based on its amino acid sequence. From this analysis a minimal putative soluble domain was selected for the purpose of cloning and expression. It was decided to clone amino acids R147 to A196 as the minimal soluble ANTAR domain. Structural predictions for the whole of the AmiR with the ANTAR domain highlighted are shown in Figure 4.3.

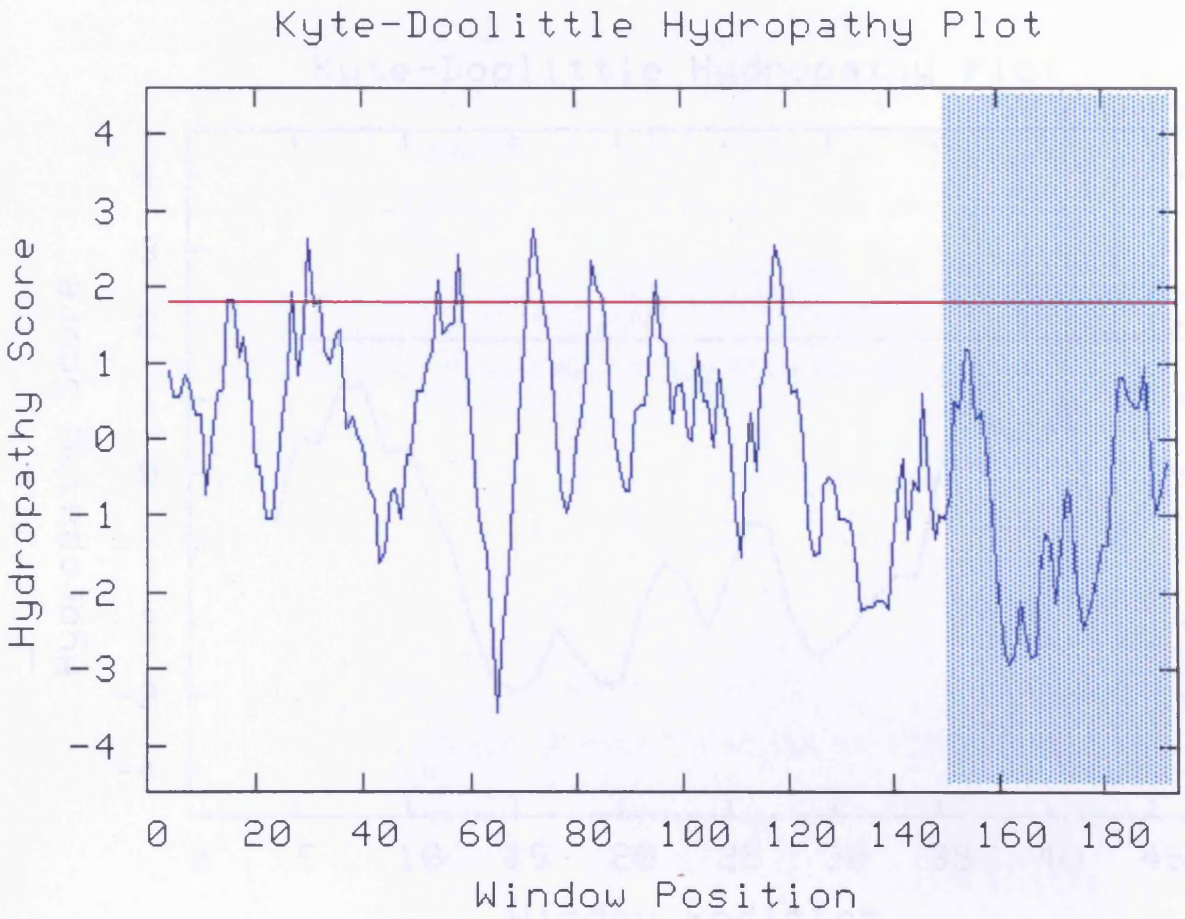


Figure 4.1. Hydropathicity plot of AmiR

The plot was generated using the Kyte-Doolittle scale [Kyte-Doolittle, 1982] and typically a value of >1.6 on the Kyte-Doolittle scale indicates a highly hydrophobic patch. The AmiR-ANTAR region is highlighted in blue.

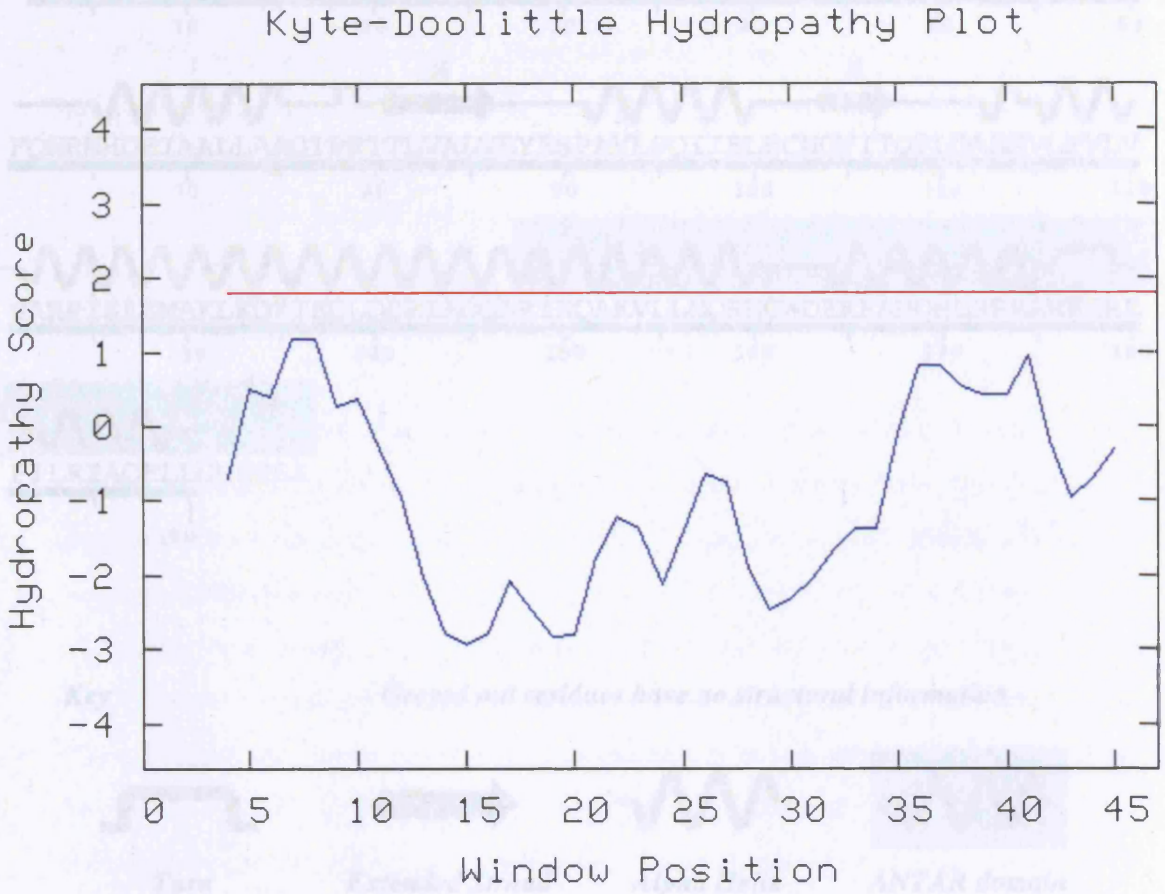


Figure 4.2. Hydropathicity plot of ANTAR.

The plot was generated using the Kyte-Doolittle scale [Kyte-Doolittle, 1982] and typically a value of >1.6 on the Kyte-Doolittle scale indicates a highly hydrophobic patch. The plot of the ANTAR region shows it is reasonably hydrophilic.

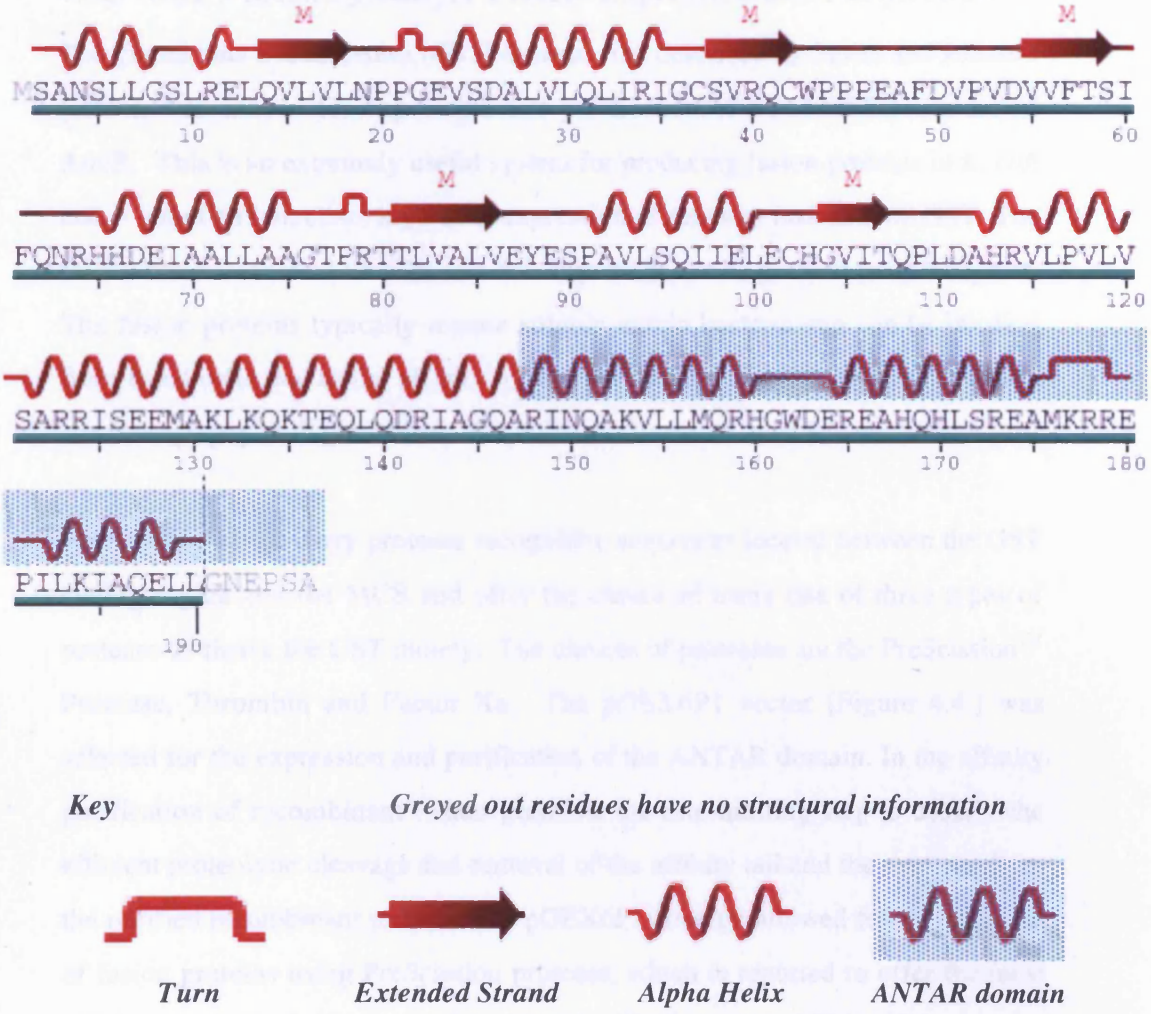


Figure 4.3. Sequence and secondary structure of AmiR

The cloned ANTAR sequence highlighted in aqua blue. Figure adapted from the PDB file 1QO0, sequence and structure analysis of the crystal structure of the AmiC-AmiR complex chain D [AmiR monomer] accessed via http://www.rcsb.org/pdb/home/home.do

4.1.3 GST Fusion System for Protein Expression and Purification

The glutathione S-transferase [GST] system, first described by Smith and Johnson [1988], was chosen for expression and purification of the ANTAR domain of AmiR. This is an extremely useful system for producing fusion proteins in *E. coli* and is based on inducible, high level expression of genes as fusions with GST. The GST protein is small, 27.5 kDa, and fusion proteins can be purified from the lysate. The fusion proteins typically remain soluble within bacteria and can be purified from lysed cells due to the affinity of the GST moiety for glutathione by affinity chromatography with glutathione immobilised to a matrix such as sepharose.

The pGEX vectors carry protease recognition sequences located between the GST coding region and the MCS and offer the choice of using one of three types of protease to cleave the GST moiety. The choices of proteases are the PreScission™ Protease, Thrombin and Factor Xa. The pGEX6P1 vector [Figure 4.4.] was selected for the expression and purification of the ANTAR domain. In the affinity purification of recombinant fusion proteins, the rate-limiting step is usually the efficient proteolytic cleavage and removal of the affinity tail and the protease from the purified recombinant protein. The pGEX6P1 strategy allowed for the cleavage of fusion proteins using PreScission protease, which is reported to offer the most efficient method of cleavage and purification [Walker *et al.*, 1994]. This protease also has the useful property of being at its most active at 4 C, thus allowing cleavage to be performed at low temperature and so improving the stability of ANTAR during purification.

4.1.3.1 PCR amplification of the ANTAR domain

Primers were designed to amplify ANTAR from pSW24 DNA using the primers *RevAntaR* and *ForAntaR* [Table 4.1]. Primer sequences were non complementary to each other and were designed to incorporate *Nde*I and *Bam*HI targets at the 5' end of the domain and an *Eco*RI and *Xho*I target at the 3' end to enable cloning into

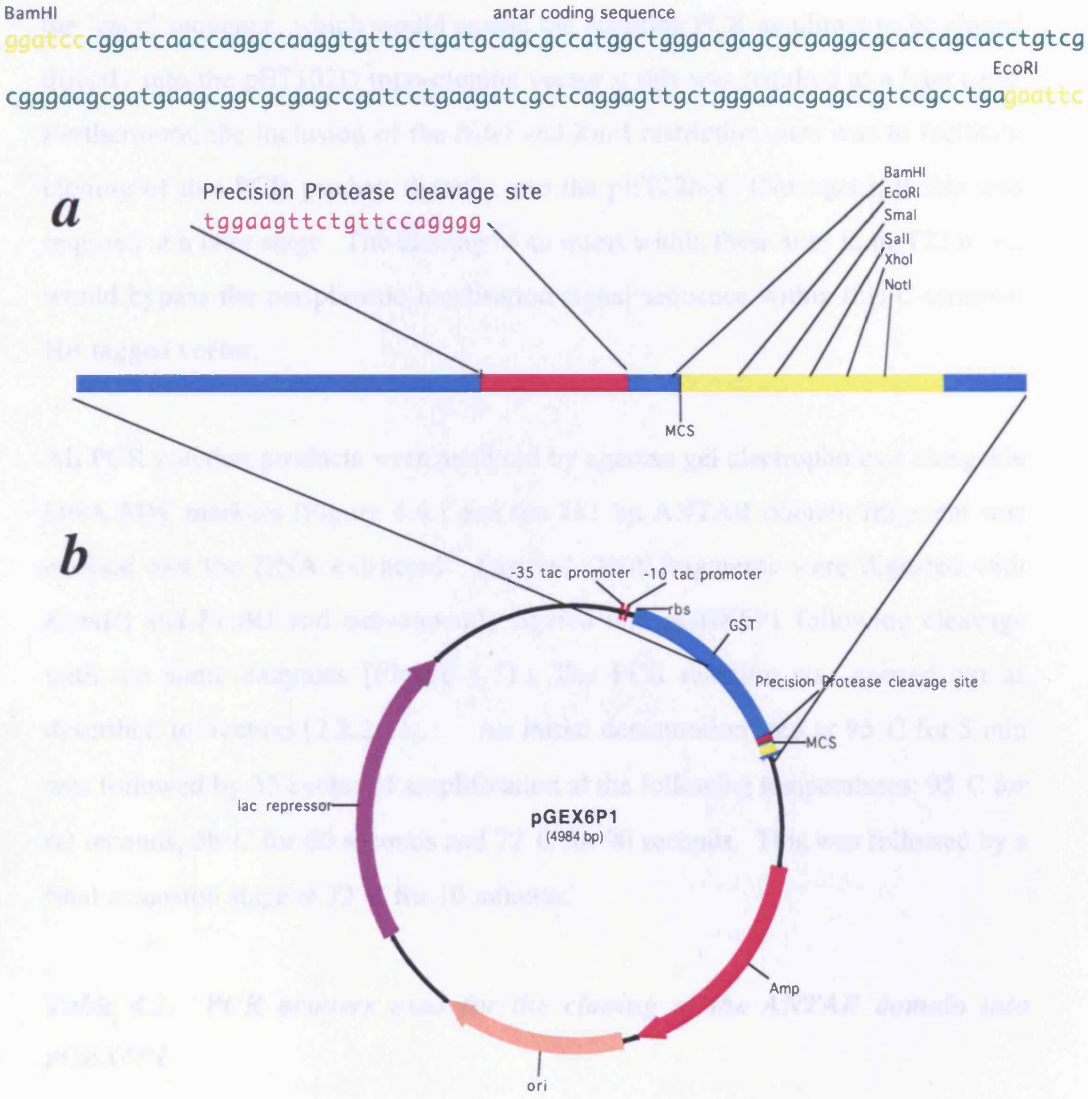


Figure 4.4. Construction of pMH8

[a] MCS of pGEX6P1 and [b] the pGEX6P1 plasmid map. The antar sequence was PCR amplified with primers which incorporated the RE sites BamHI and EcoRI, which allowed cloning into these sites within the MCS of pGEX6P1.

either the pGEX6P1 or pET22b(+) vectors. In addition, the primers also contained the 'cacc' sequence, which would enable the resulting PCR amplimer to be cloned directly into the pET102D topo-cloning vector if this was required at a later time. Furthermore, the inclusion of the *Nde*I and *Xho*I restriction sites was to facilitate cloning of this PCR product directly into the pET22b(+) [Novagen], if this was required at a later stage. The cloning of an insert within these sites in pET22 b (+), would bypass the periplasmic localisation signal sequence within this C-terminal His tagged vector.

All PCR reaction products were analysed by agarose gel electrophoresis alongside DNA MW markers [Figure 4.4.] and the 181 bp ANTAR domain fragment was excised and the DNA extracted. Purified DNA fragments were digested with *Bam*HI and *Eco*RI and subsequently ligated into pGEX6P1 following cleavage with the same enzymes [Figure 4.5] . The PCR reaction was carried out as described in Section [2.2.2.11]. An initial denaturation step at 95 C for 5 min was followed by 35 cycles of amplification at the following temperatures: 95 C for 60 seconds, 58 C for 60 seconds and 72 C for 90 seconds. This was followed by a final extension stage at 72 C for 10 minutes.

Table 4.1. PCR primers used for the cloning of the ANTAR domain into pGEX6P1

Name	*Sequence [5' –3']*	*Features*
ForAntar	CACCCATATGGGATCCCGGATCAACCAGGCCAAG	NdeI BamHI
RevAntar	GGTAATTAACTCGAGGAATTCTCAGGCGGACGGC TCGTTTCCCAGC	*Xho*I *Eco*RI

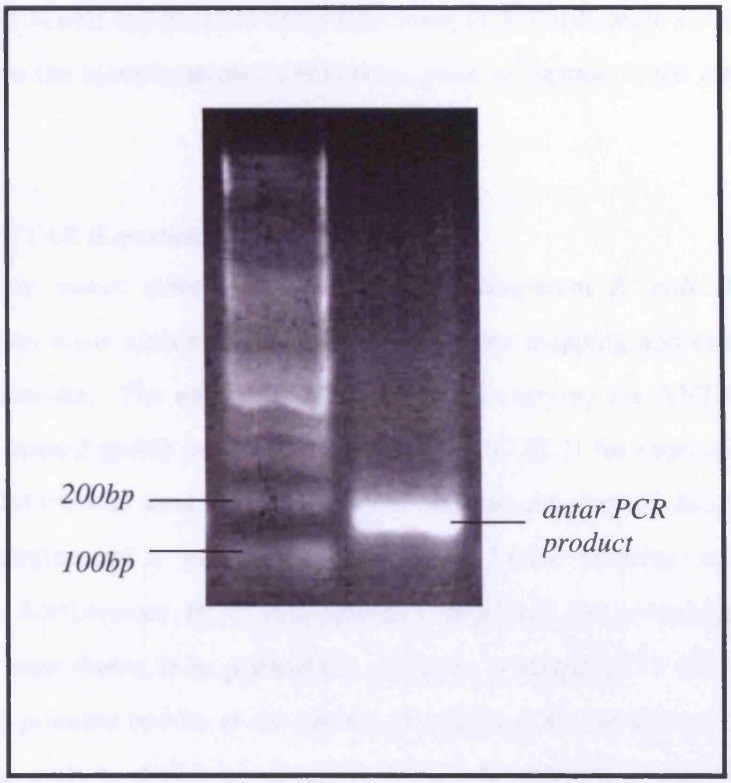


Figure 4.5. The ANTAR PCR product

The PCR product was visualised on a 1.5% agarose Gel run alongside a 2-log DNA ladder. The ANTAR PCR product was subsequently ligated into pGEX6P1, when both insert and vector had been digested with BamHI and EcoRI to create ANTAR expression vector pMH8.

4.1.3.2 Cloning of ANTAR PCR product into pGEX6P1 protein expression vector to Create pMH8

For cloning of DNA fragments into pGEX6P1, vector DNA was digested with *Bam*H1 and *Eco*RI and purified using QIAquick PCR Purification Kit [QIAGEN] according to the manufacturers' instructions, prior to ligation to the cleaved PCR product.

4.1.3.3 ANTAR Expression in BL21

The ligation mixes were used to transform competent *E. coli* JM105 and recombinants were identified by restriction enzyme mapping and confirmed by DNA sequencing. The new recombinant plasmid carrying the ANTAR domain was newly named pMH8 and transformed into *E. coli* BL21 for expression studies. Although JM105 was used as a host strain to maintain the plasmid, BL21, which is specially engineered to maximise expression of fusion proteins, was used for expression. Furthermore, BL21 is deficient in *OmpT* and *Lon* protease production, which has been shown to be particularly useful for producing GST fusion proteins. The OmpT protease resides at the surface of wild type *E. coli* strains, presumably helping the cells to derive amino acids from their external environment. Cells deficient in both these proteases are much more amenable to the production of proteins from cloned genes [Grodberg and Dunn, 1988].

4.1.4 Expression and Purification of GST-ANTAR

Expression and purification of the ANTAR domain from GST sepharose was carried as described in Section 2.2.3. The AmiR ANTAR was isolated from IPTG induced BL21 [DE3], pLysS, pMH8 cells. The final cell pellet was lysed by sonication and cleared by centrifugation [Section 2.2.3.10]. Recombinant GST-ANTAR was isolated bound to the GST-sepharose and the AmiR ANTAR separated from the fusion tag using PreScission Protease as [Sections 2.2.3.11 and 2.2.3.12], and isolated using Gel Filtration Chromatography [Section 2.2.4].

4.1.5 Gel Filtration Chromatography

Following purification of GST-fusion proteins from other non GST-sepharose binding proteins, the GST moiety was cleaved using PreScission protease as described in Section 2.2.3.12. Elution of the cleaved recombinant protein and subsequent analysis on SDS-PAGE shows a number of contaminant bands that co-eluted with the desired protein fragment [Figure 4.6 (a)]. As these contaminants were of different molecular weights, they were removed using gel filtration chromatography.

Gel filtration chromatography is routinely used to separate proteins, peptides, and oligonucleotides on the basis of size. This technique was first described in 1965 for the separation of phenol and RNA using Sephadex G-25, and this form of chromatography is based on the observation that as molecules migrate through a bed of porous beads, they diffuse into the beads with varying degrees. In the gel filtration column, sample molecules are partitioned between the eluent [mobile phase], and the penetrable pores of the beads [stationary phase]. This partitioning creates a dynamic equilibrium of sample molecules between the mobile and the stationary phases and is driven exclusively by diffusion. Smaller molecules diffuse deeper within the pores of the beads thus moving through the bed more slowly, while larger molecules diffuse slower into the beads or not at all, and hence exit the bed of resin more quickly. Thus, the shorter the retention time of the molecule within the bed of resin, the larger the molecular weight and/or three dimensional shape of the molecule.

SDS-PAGE of the ANTAR protein following Gel Filtration can be seen in Figure 4.6 (b) and the Gel Filtration elution profile of ANTAR can be seen in Figure 4.7.

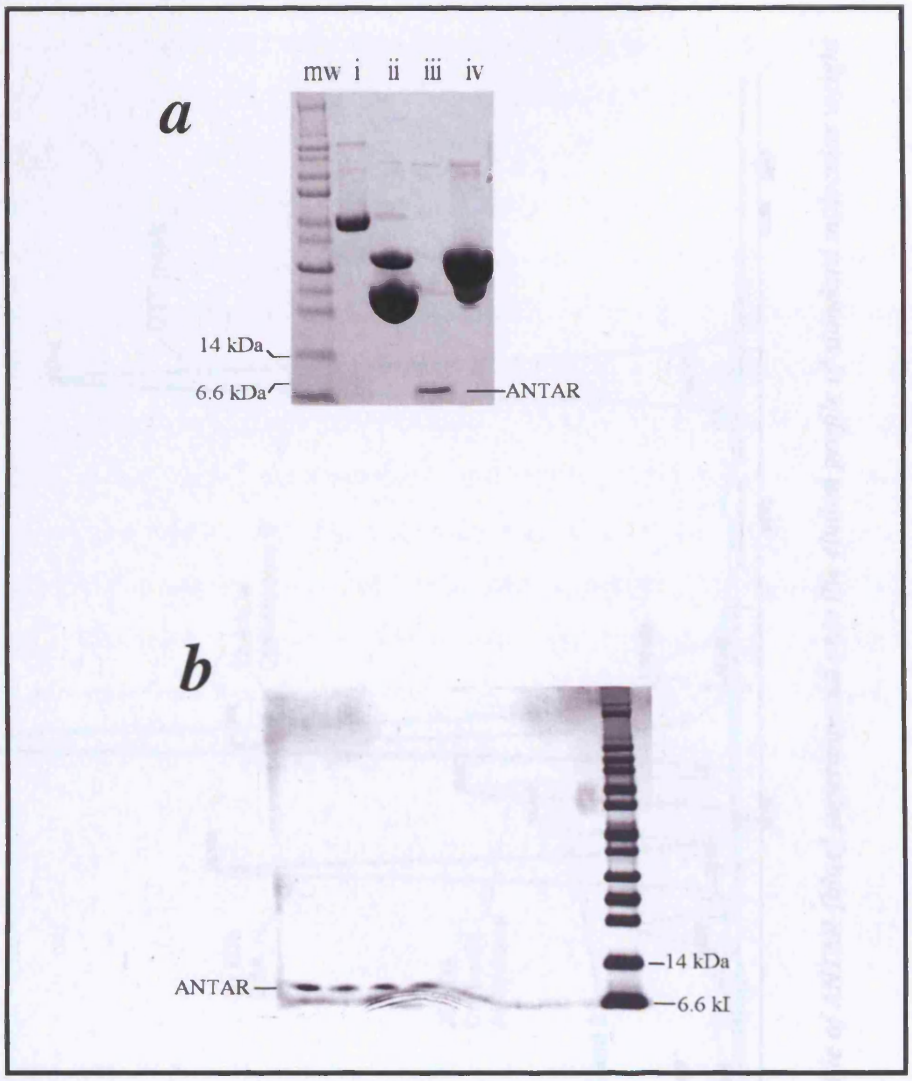


Figure 4.6. Purification of ANTAR.

[a] Gel showing purification of GST-fusion ANTAR protein. Lane [i] the large band at approximately 40 kDa is the PreScission™ protease. [ii] GST-fusion ANTAR following cleavage with PreScission™ protease. [iii] Eluate of the cleaved ANTAR protein. [iv] fusion band of approximately 35 kDa [the m.w. of the ANTAR protein fused to GST] is seen, hence the chosen construct is highly soluble. [b] ANTAR following the removal of contaminants by gel filtration.

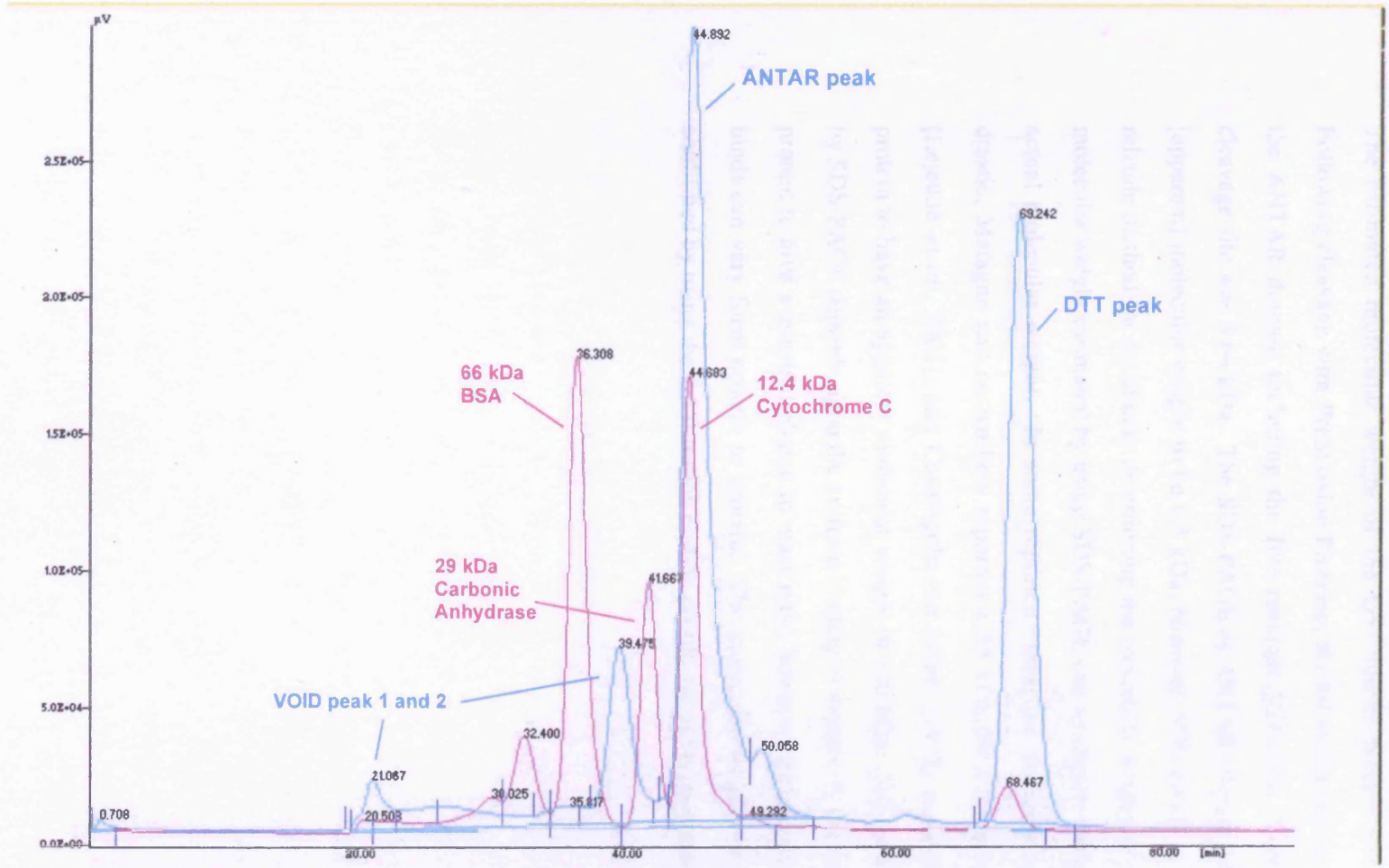


Figure 4.7. Gel filtration elution profile of ANTAR [blue] superimposed onto the elution profile of standard molecular weight markers [pink].

The estimated molecular weight of the GST-fusion protein was 35.09 kDa. Following cleavage with PreScission Protease, the estimated molecular weight of the ANTAR domain including the five residues [*GPLGS*], C-terminal to the cleavage site was 5.64 kDa. The SDS-PAGE of ANTAR [Figure 4.6], shows the [apparent] molecular weight to be 6.5 kDa, however, SDS-PAGE is not always a reliable method for accurately determining the molecular weight of proteins. The molecular weight estimated by using SDS-PAGE can be slightly different from the actual molecular weight. In some reported cases, the difference can be quite drastic, Matagne and co-workers reported a 55 kDa for a 29 kDa β-lactamase [Lejeune *et al.*, 2001], and Casaregola and others [1992] reported a 114 kDa protein to have an apparent molecular weight of 180 kDa. Separation of proteins by SDS-PAGE depends upon the uniform binding of negatively charged SDS to the protein to give a constant charge to mass ratio, however, the amount of SDS that binds can vary form protein to protein. The molecular weight for ANTAR was confirmed by using the gel filtration elution profile and mass spectroscopy.

4.2 Characterisation of the ANTAR Domain

4.2.1 Introduction

Following elution of the ANTAR domain from the GST column, a number of higher molecular weight contaminating bands could still be observed. In order to purify the ANTAR domain from the contaminants, gel filtration chromatography was carried out as described in Section 2.2.4. Calibration experiments where molecular weight standards were run through the S200 gel filtration column [Figure 4.7], showed the 66 kDa marker [BSA] eluted after approximately 36 minutes, the 29 kDa marker [carbonic anhydrase] eluted after approximately 41 minutes and the 12.4 kDa marker [Cytochrome C] eluted at approximately 44.6 minutes. The ANTAR protein in this experiment eluted at approximately 44.8 minutes following injection of sample indicating a species with a molecular weight of slightly greater than 12.4. Thus, the ANTAR molecule seems to self-associate. Dynamic Light Scattering investigated whether this was due to aggregation of the protein.

4.2.2 Mass Spectrometry Sequencing of ANTAR

Prior to the commencement of any analytical work with the purified protein sample, Mass Spectrometry sequencing of the protein was used to confirm that the sample was indeed ANTAR and to evaluate the purity of the sample. Mass spectrometry can give sequence data from any Coomassie stained band or gel spot [Scheler, *et al.*, 1998]. The protein is digested with trypsin in the gel [Section 2.2.6.1], peptides eluted and fractionated by reverse phase chromatography and introduced into the mass spectrometer. The mass spectrometer determines the mass of the peptides and the sequence. Mass Spec reactions were carried out by Leon G. D'Cruz. Division of Protein Structure, National Institute of Medical Research, where a research grade MALDI-TOF is used, coupled with SDS-PAGE, to identify proteins from their peptide mass fingerprints. For the analysis, a trypsin digested ANTAR, which had been run on a SDS-PAGE was used.

4.2.3 Theory of Dynamic Light Scattering [DLS]

Dynamic Light Scattering is a technique which provides information about the size and distribution of macromolecules in solution and allows for determination of the hydrodynamic radius, translational diffusion coefficient, polydispersity, and an estimated molecular weight of a macromolecule [Ferre-D'Amare and Burley, 1997].

DLS is frequently used for the determination of the homogeneity of a protein sample prior to crystallization trials, estimation of the molecular weight of a macromolecule, and protein-ligand interaction studies. It is often used in the examination of conformational changes and self-association properties of macromolecules as a function of pH, ionic strength, temperature, and other solvent conditions. DLS of the purified ANTAR domain was carried out to assess the purity of the sample and to survey the solubility of the sample through the interpretation of the polydisperisty of the sample. DLS of ANTAR was also used to confirm the oligomeric state of ANTAR; the Gel Elution Profile for ANTAR indicated that ANTAR was dimeric, this was confirmed using DLS,

4.2.3.1 Interpretation of the DLS measurements

The determination of the extent of polydispersity in a sample is carried out as follows. The average measured radius of polydispersed particles [polyD (nm)], is divided by the average hydrodynamic radius [RH], multiplied by a 100 to give the percentage polydispersity [Figure 4.8 (C)]. Interpretation of the data is summarised in Tables 4.2, 4.3 and 4.4.

Table 4.2. Interpretation of percentage polydispersity

Percentage polydispersity	*Interpretation*
< 15 %	Negligible polydispersity
< 30 %	Moderate amount of polydispersity
> 30 %	Significant amount of polydispersity

The baseline values determined are interpreted as follows:

Table 4.3. Interpretation of baseline values

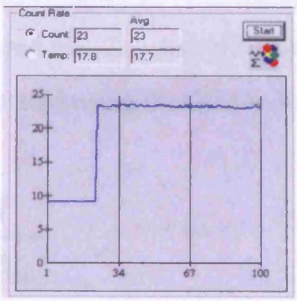
Baseline	*Interpretation*
0.997-1.001	Narrow monomodal size distribution
1.002-1.005	Broad monomodal size distribution
> 1.005	Size distribution cannot be resolved

The Sum-of-squares [SOS] values, which correlate to the background noise levels, are interpreted as follows:

Table 4.4. Interpretation of SOS values

Sum of Squares Value	*Interpretation*
1.000-5.000	Negligible error, low background noise
5.000-20.000	Somewhat significant background noise
> 20.000	Very high error and/or background noise

a



b

M:ntl	Time(s)	Temp(C)	Count Rate	Ampl	Diff Coeff	Radius(nm)	Polyd(nm)	Polydindx	MW(KDa)	%Mass	Baseline	Sos Error
1	0.0	17.4	109119	0.518	967	1.86	1.23	0.44	14.4	100.0	1.001	14.2
2	8.0	17.7	101955	0.502	1.14e+003	1.60	0.949	0.35	10.1	100.0	0.997	8.71
3	16.0	17.9	102448	0.508	1.12e+003	1.63	0.631	0.15	10.5	100.0	1.001	9.41
4	24.0	17.7	101098	0.501	1.13e+003	1.60	1.01	0.40	10.1	100.0	1.000	10.6
5	31.0	17.9	101687	0.511	1.10e+003	1.66	0.895	0.29	11.0	100.0	1.001	9.80
6	38.0	17.7	103775	0.508	1.10e+003	1.65	1.01	0.37	10.8	100.0	1.006	10.6
7	46.0	17.7	105360	0.515	1.08e+003	1.69	0.824	0.24	11.4	100.0	0.996	9.27
8	53.0	17.8	102298	0.515	1.13e+003	1.61	0.658	0.17	10.3	100.0	0.995	9.10
9	61.0	17.9	102327	0.517	1.11e+003	1.65	0.834	0.26	10.8	100.0	0.997	12.3
10	68.0	17.7	102072	0.519	1.15e+003	1.58	0.986	0.39	9.78	100.0	0.998	11.0
Aves:												
Mono		17.7	103213	0.511	1.10e+003	1.65	0.903	0.31	10.9	100.0	0.999	10.5
B=1		17.7	103213	0.267	2.51e+003	0.753	---	---	1.85	97.3	0.999	1.38
B=2				0.297	717	2.59			32.7	2.7		

c

$$\frac{0.903}{1.65} \times 100\% = 54.7\ \%$$

d

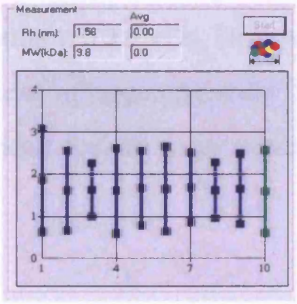


Figure 4.8. Dynamic Light Scattering of the ANTAR protein.

[a] The average count rate of water was taken prior to analysing the ANTAR protein sample. The count rates of pure clean water was obtained at 23 counts, with negligible fluctuation. [b] DLS data on the ANTAR protein. [c] calculation of polydispersity. [d] Screen shot highlighting one of the 10 measurements.

4.2.3.2 The Stokes-Einstein relationship

From the measured translational diffusion coefficient DT, the hydrodynamic radius RH can be calculated using the Stokes-Einstein relation

$$D_T = \frac{k_B T}{6\pi\eta R_H}$$

with the Boltzmann constant kB, the temperature T in Kelvin and h being the viscosity of the solvent. Molecular masses are estimated from RH using an empirically derived relationship between RH and molecular masses for a number of well-characterised globular proteins. Values for DT reported are statistical averages over at least 10 independent measurements.

4.2.3.3 Dynamic Light Scattering analysis of the ANTAR protein

From the data in figure 4.8 [b], the average molecular weight of the ANTAR protein in solution is 10.9 kDa. However, the cloned ANTAR gene would yield a protein fragment of approximately 5.6 kDa. The SOS values and baseline parameters indicate that the noise values were relatively low.

4.2.4 Summary of Chapter 4

It appears as if ANTAR is behaving as a dimer in solution [5.6 x 2 = 11.2 kDa] as the calculated value of the dimer is close to the observed DLS data of 10.9 kDa. However, some aggregation could be taking place since the polydispersity is 54.7% with relation to the hydrodynamic radius. This could account for the gel-filtration elution profile of the ANTAR protein. There are no free cysteines in the cloned ANTAR protein sequence, thus intermolecular disulphide bonding is unlikely to be participating in the oligomerisation states of ANTAR. It is unknown if the oligomerisation of ANTAR is a natural phenomenon or if it could perhaps serve some functional role *in vivo*. The presence of the long α-helical stem in AmiR could perhaps be preventing the formation of multimeric states of the ANTAR domain.

5 *Analysis of the ANTAR-RNA Interaction*

5.1 *Identification of Minimal Leader RNA Transcript Studies Using Site Directed Mutagenesis*

5.1.1 *Introduction*

AmiR, NasT and NasR share carboxyl-terminal sequence similarity that defines a family of transcription antiterminator proteins resulting in the classification of the ANTAR domain [Shu and Zhulin 2002]. Due to the structural similarity of the antiterminator proteins of the *ami* and *nas* operons, the possibility exists that the mechanism of action will be similar for the two proteins, although this remains to be determined. Sequence analysis of the leader regions of the *nas* operon of *K. oxytoca* M5aI has revealed some similarities between the two [Chai and Stewart 1999].

The upstream leader region of the *nas* operon contains a potential stem loop structure called 1:2, followed by a second terminator stem loop structure called 3:4. The upstream *amiE* leader region also contains a potential 1:2 structure and a rho-independent 3:4 terminator sequence, similar to that found in the *nas* operon leader sequence. Both examples of the rho-independent terminators of the *ami* and *nas* operons fold into a highly stable stem loop structures, having the characteristics of a transcription terminator: A G+C rich stem loop, capped by a stabilizing GNRA tetra-loop [Heus and Pardi, 1991]. GNRA tetraloops are said to play a major role in large protein-RNA complexes such as the ribosome and are frequently found in large RNA folds because of their conformational stability and their ability to engage in tertiary contacts with other RNA motifs [Correll and Swinger, 2003;

Herman and Westhof, 1999]. The *ami* and *nas* terminator stem loops are immediately followed by a polythymidine tract of nine residues, also characteristic of terminator structures. Deletions to the 3:4 stem loop in the *nasF* operon lead to constitutive operon expression [Chai and Stewart 1999]. It has previously been shown that disruption to the 3:4 stemloop of the *ami* operon leads to constitutive amidase operon expression [Wilson *et al.*, 1996]. Mutagenesis and deletions to regions within the leader RNA of the amidase operon have identified some regions within the leader transcript that are essential for antitermination and thus AmiR binding [residues 36-41 and 54-62] [Drew and Lowe, 1989]. Binding by AmiR to this region would be expected to result in disruption to the formation of the *rho*-independent terminator loop allowing transcription to proceed to the downstream genes. Thus, regulation of operon expression must be exerted via this antiterminator structure, - an AmiR/RNA complex which prevents formation of the RNA terminator structure.

Alterations disrupting the base-pairing in the upper portion of the 1:2 stem abolishes antitermination and NasR binding in the *nas* operon. When the uninducibility and binding defects were restored by compensatory changes that restored pairing, the defects were repaired. These observations led to the theory that the 1:2 stem-loop contributed to the NasR binding site [Chai and Stewart 1999]. The equivalent 1:2 stem-loop structures of the *ami* and *nas* operons are capped with a 6 nt-loop with adenine and guanine as the first and fourth residues respectively. Chai and Stewart [1999] showed that for *nas* operon expression, these residues are critical for antitermination regulation. Functions of the 1:2 structures in the *ami* and *nas* operons are presently unknown, and one hypothesis is that they may form a part of a network of alternative secondary structures whose formations are controlled by the antiterminator protein.

5.1.2 *Investigating the leader region of the amidase operon using leader mutants and leader deleted constructs*

To identify the minimal RNA species required for the antitermination reaction and to investigate the role of the first stem loop structure found within the *ami* leader sequence, a series of leader mutations and deletions were constructed. These were carried out for two reasons; prior to experimental studies into the interaction between ANTAR and the *ami* leader sequence using *in vitro* transcribed RNA, it was necessary to scale up the production of the RNA leader sequence using an *in vitro* transcription kit, and to further investigate the function of the *ami* operon 1:2 stem-loop

5.1.2.1 Construction of pMH1

The leader sequence of the amidase operon was sub-cloned into pGEM4Z to enable the future production of *in vitro* transcribed run off transcripts. The amidase operon leader sequence and *amiE* gene *Hin*dIII-*Kpn*1 fragment, was sub-cloned from pJB950, after separating the small *amiE* gene fragment [2.4 kb] from the larger fragment by agarose gel electrophoresis and the fragment ligated into the *Hin*dIII-*Kpn*I cut pGEM4Z [Figure 5.1], in an orientation such that *amiE* can be expressed from its own or the vector T7 and/or *lac* promoters, to create pMH1. A representational map of pMH1 is shown in Figure 5.2, and the DNA sequence of the pMH1 *ami* leader region in Figure 5.3.

5.1.2.2 Construction of pMH2 and pMH3

Using *in vitro* site directed mutagenesis, an additional *Hin*dIII site was introduced into pMH1 operon leader region 118 bp downstream of the first *Hin*dIII site creating the construct pMH2. This changed the DNA sequences immediately downstream of the *ami* promoter –10 region and before the transcription start site. The 118 bp region flanked by the two *Hin*dIII sites was excised and subsequent

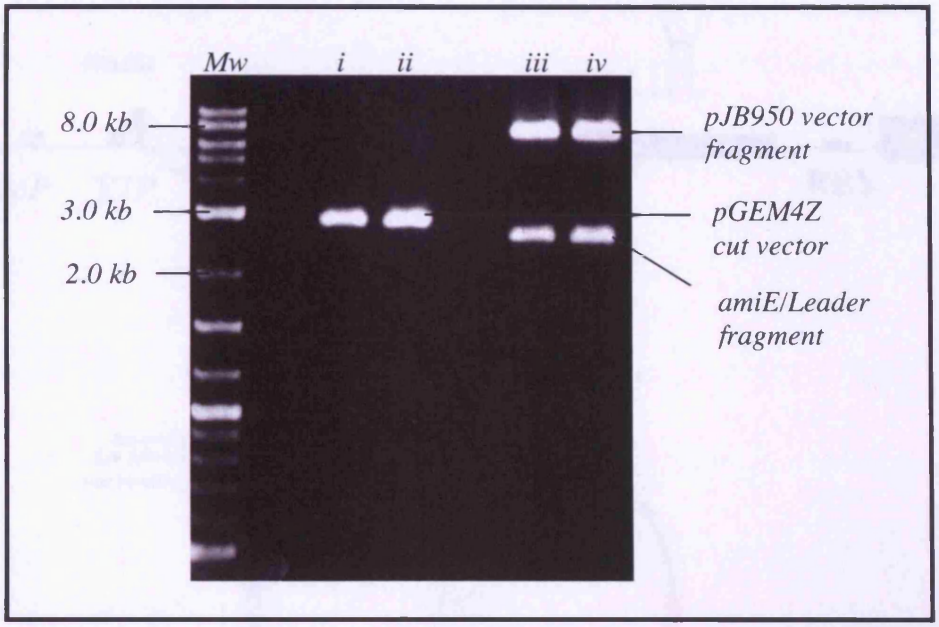


Figure 5.1. Cloning the ami leader fragment from pJB950 into pGEM4Z

1.2% Agarose gel electrophoresis of HindIII/KpnI digestied [i and ii], 2.9kb pGEM4Z vector. In lanes iii and iv, contain the HindIII-KpnI cut pJB950 vector with the 2.4kb liberated amiE/Leader fragment involved in the construction of pMH1.

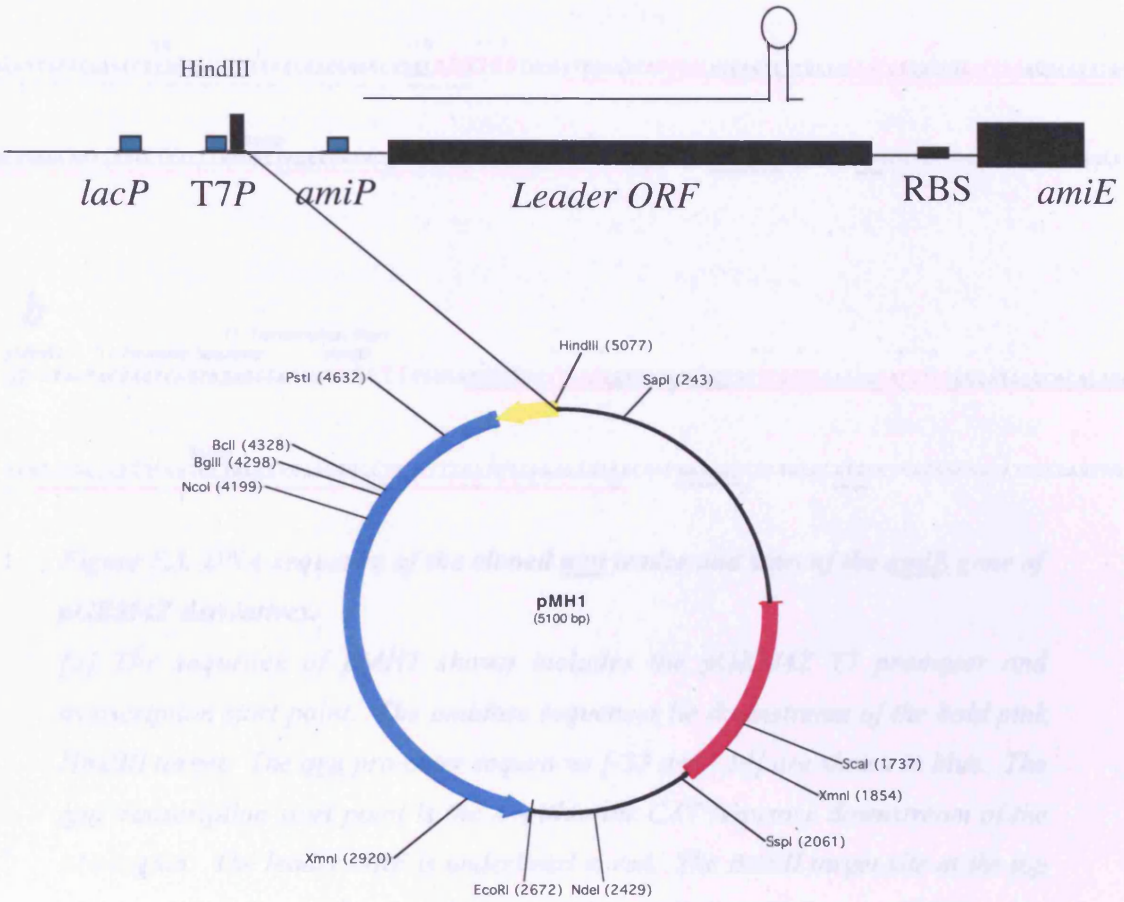


Figure 5.2. Plasmid map of pMH1

The map shows the ami leader region [yellow] and amiE gene [aqua blue] and some restriction enzyme sites. The Leader sequence is represented schematically presented above the construct and shows the locations of the lac, T7 and ami promoters.

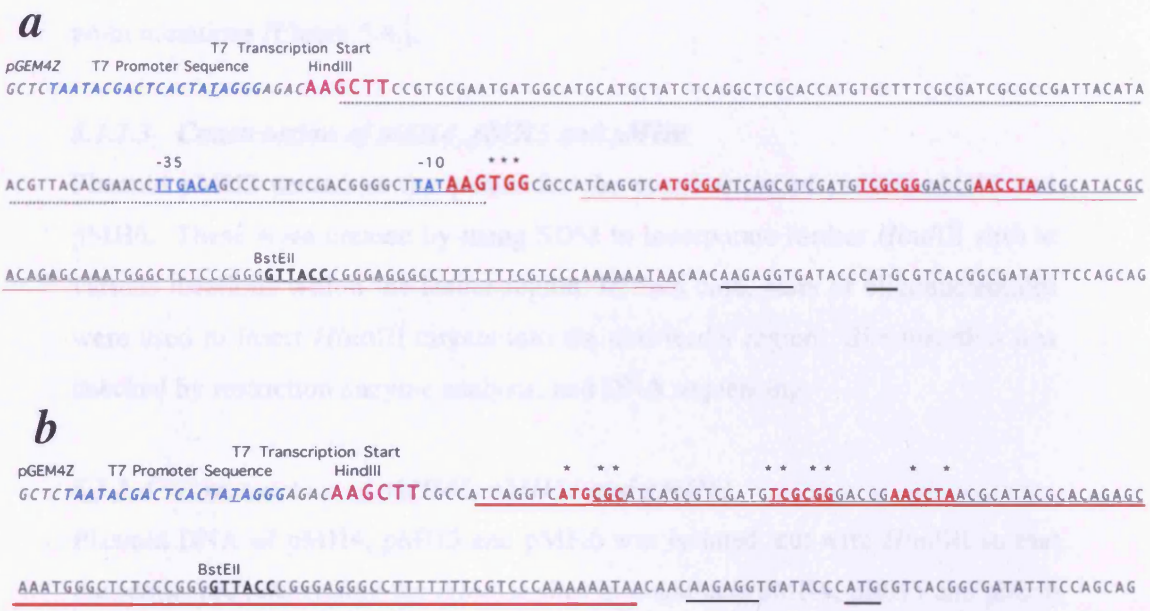


Figure 5.3. DNA sequence of the cloned ami leader and start of the amiE gene of pGEM4Z derivatives.

[a] The sequence of pMH1 shown includes the pGEM4Z T7 promoter and transcription start point. The amidase sequences lie downstream of the bold pink HindIII target. The ami promoter sequences [-35 and –10] are shown in blue. The ami transcription start point is the A within the CAT sequence downstream of the –10 region. The leader ORF is underlined in red. The BstEII target site at the top of the terminator stem/loop is shown and the amiE Shine/Dalgarno [S/D] region and start codon are underlined in black. To create pMH2 the bold red AAGTGG sequence was mutated to AAGCTT to create a second HindIII target. To create pMH3 the DNA sequences between the two HindIII targets were deleted. [b] The sequence of the pMH3 ami leader region. The vector T7 promoter and transcription start point are shown. The HindIII site joining the vector and ami sequence is shown in bold pink [AAGCTT]. The leader ORF is underlined in red. The BstEII target is shown and the amiE S/D and start cds are underlined in black. Sequences shown in red were changed to HindIII targets to create pMH4 [ATCCGC], pMH5 [TCGCGG] and pMH6 [AACCTA].

ligation of the *Hind*III sticky ends created pMH3, the parent plasmid for further point mutations [Figure 5.4.].

5.1.2.3 Construction of pMH4, pMH5 and pMH6

Plasmid pMH3 served as the parent for the construction of pMH4, pMH5 and pMH6. These were created by using SDM to incorporate further *Hind*III sites at various locations within the leader region. In each case, pairs of oligonucleotides were used to insert *Hin*dIII targets into the *ami* leader region. Site insertion was checked by restriction enzyme analysis, and DNA sequencing.

5.1.2.4 Construction of pMH41, pMH51 and pMH61

Plasmid DNA of pMH4, pMH5 and pMH6 was isolated, cut with *Hin*dIII so that the sequences intervening the *Hin*dIII sites in constructs pMH4, pMH5 and pMH6 were subsequently deleted creating pMH41, pMH51 and pMH61 respectively [Figures 5.5., 5.6. and 5.7.].

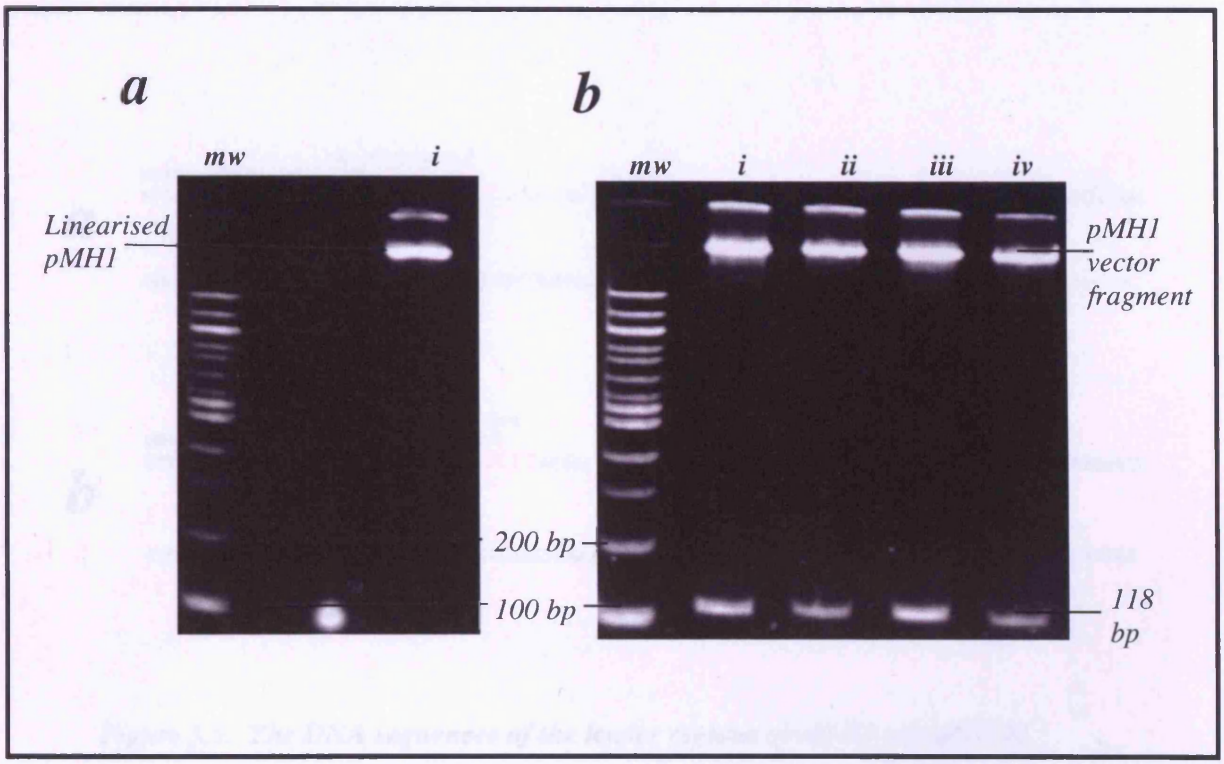


Figure 5.4. Construction of pMH2 and pMH3.

Polyacrylamide Gel Electrophoresis of [a] HindIII digested pMH1 DNA [Lane i], and [b] Lanes i-iv showing HindIII digested DNA from potential pMH2 clones, run alongside a 2-log Mw DNA ladder. The mutation in pMH1 created an additional HindIII site resulting in release of a 118 bp fragment upon HindIII digestion.

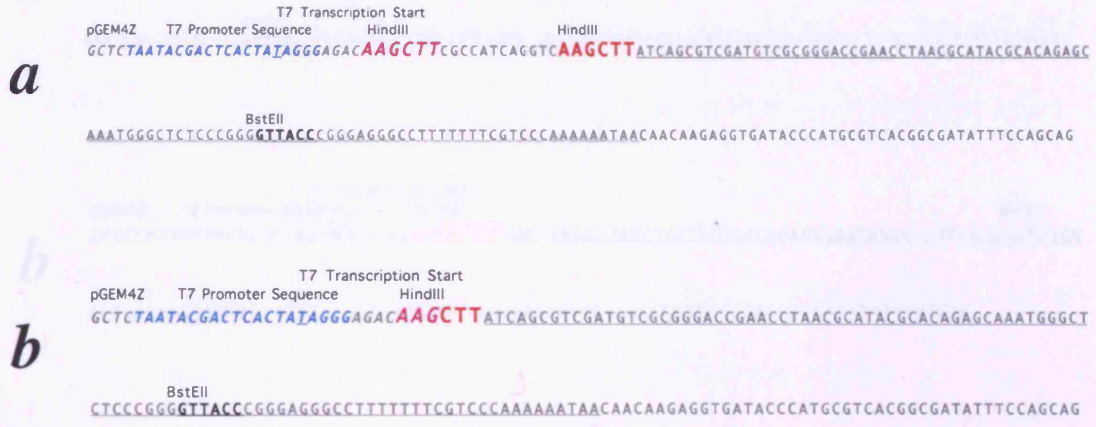


Figure 5.5. The DNA sequences of the leader regions of pMH4 and pMH41

pGEM4Z Vector sequences are italicised and the T7 promoter region is in aqua blue. [a] pMH4 with the WT HindIII site and the nascent HindIII site shown in red [b] pMH41 formed by deletion and joining of the HindIII sites of pMH4.

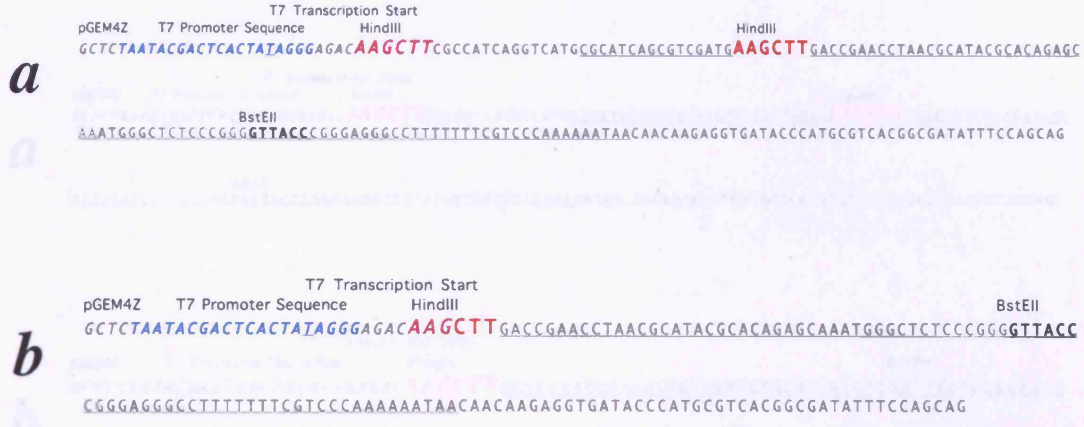


Figure 5.6. The DNA sequences of the leader regions of pMH5 and pMH51

pGEM4Z Vector sequences are italicised and the T7 promoter region is in aqua blue. [a] pMH5 with the WT HindIII site and the nascent HindIII site shown in red [b] pMH51 formed by deletion and joining of the HindIII sites of pMH5.

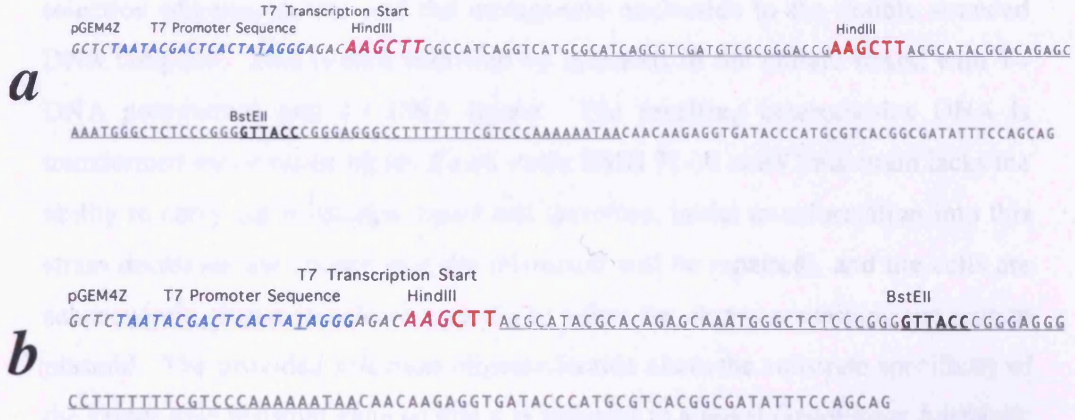


Figure 5.7. The DNA sequences of the leader regions of pMH6 and pMH61

pGEM4Z Vector sequences are italicised and the T7 promoter region is in aqua blue. [a] pMH6 with the WT HindIII site and the nascent HindIII site shown in red [b] pMH61 formed by deletion and joining of the HindIII sites of pMH6.

5.1.3 Methods for the Incorporation of Site-Directed Mutations into the ami Leader sequence.

Two different methods were employed to obtain the mutants created by SDM, the GeneEditor™ *in vitro* Site Directed Mutagenesis System, and the Stratagene QuickChange system. The former system involves annealing of a provided selection oligonucleotide and the mutageneic nucleotide to the double stranded DNA template. This is then followed by synthesis of the mutant strand with T4 DNA polymerase and T4 DNA ligase. The resulting heteroduplex DNA is transformed into a repair minus *E.coli* strain BMH 71-18 *mutS* [this strain lacks the ability to carry out mismatch repair and therefore, initial transformation into this strain decreases the chance that the mismatch will be repaired], and the cells are subsequently grown in selective media to select for clones containing the mutant plasmid. The provided selection oligonucleotide alters the substrate specificity of the vector *amp* resistant gene so that it is resistant to a novel GeneEditor Antibiotic Selection Mix. Plasmids resistant to this were subsequently isolated and transformed into the final host strain, JM109 for analysis.

This system was superseded by the modified QuickChange method due to the fact that once the initial site directed mutagenesis reaction was carried out, subsequent mutagenesis reactions could not be carried out with the same plasmid bearing the first mutation as the selection procedure was made redundant by the altered *amp* resistance gene changed in the first round.

The Stratagene QuickChange SDM kit was developed by Vandeyar and co-workers [1988] and the method is described in Section [2.2.2.14] of this thesis. The pGEM4Z based pMH3, which contains the WT leader sequence and *amiE* gene was used as the template DNA. Two synthetic oligonucleotides primers were designed for each desired mutation [Table 5.1]. .

Table 5.1. Primers used for the SDM of ami-Leader sequence

Name	*Sequence [5' –3'] HindIII Site in Bold and point mutation in red*
MH2F	[Phos]GCATGACCTGATGGCGAAGCTTATAAGCCCCGTCGG
MH2R	Reverse Compliment of above sequence
MH4F	[Phos]CGACATCGACGCTGATAAGCTTGACCTGATGGCGAAGC
MH4R	Reverse Compliment of above sequence
MH5F	[Phos]CGTATGCGTTAGGTTCGGTCAAGCTTCATCGACGCTGATGCGC
MH5R	Reverse Compliment of above sequence
MH6F	[Phos]GCTCTGTGCGTATGCGTAAGCTTCGGTCCCGCGACATCG
MH6R	Reverse Compliment of above sequence

5.1.4 Analysis of Mutant Leader Regions

To assess the effects of the mutations and deletions to the leader RNA on the antitermination reaction, the constructs were assayed for amidase activity. Amidase assays were performed in JM109 carrying these constructs, in the presence and absence of AmiR [pSW35 SM^R]. Analysis in the absence of pSW35 would show if the changes had affected the normal termination reaction and in the presence of pSW35 would be a measure of the antitermination ability of AmiR to interact with the mutated/shortened leader RNA.

The system is further complicated by the presence of two functional upstream promoters. Expression of amidase in pMH1 and pMH2 can occur following transcription initiation from both the vector *lac* and *ami* promoters. Expression via antitermination from pMH3 and all subsequent constructs derive only from transcription initiation at the *lac* promoter since the amidase promoter is missing in these constructs.

5.1.4.1 Analysis of pMH1 and pMH2

Plasmid pMH1 acts as a positive control in these experiments. Amidase can be produced from transcription originating at the amidase operon promoter and the

vector *lac* promoter. In the absence of AmiR pMH1 shows background amidase activity levels of 1.5, increased four-fold [to 5.7] in the presence of AmiR [pSW35]. The insertion into pMH1 of the *Hind*III target immediately downstream of the amidase promoter –10 region [pMH2] caused slight [1.4-fold] increase in background [AmiR independent] amidase activity but no change to the AmiR-dependent level of amidase activity [Figure 5.8].

5.1.4.2 Analysis of pMH3, pMH4, pMH5 and pMH6

Amidase expression from these constructs derives from transcription from the *lac* promoter. It is apparent from the assay results for all of the constructs that high levels of amidase activity are present even in the absence of AmiR. These all produce full length *ami* leader transcripts fused to the vector sequences from the vector *lac* promoter. In the absence of AmiR, pMH3 produces an amidase activity of 10.5 and the average values for pMH3, pMH4, pMH5 and pMH6 is 9.4. It would appear that the relatively high values are due to the strength of the lac promoter and the presence of additional upstream RNA sequences folding in such a way as to interfere with formation of the termination loop.

Another possibility is that the *P. aeruginosa* sequences missing in these constructs [from *Hin*dIII to the *ami* promoter] have a role in down regulating expression from the *ami* promoter. This was not further examined.

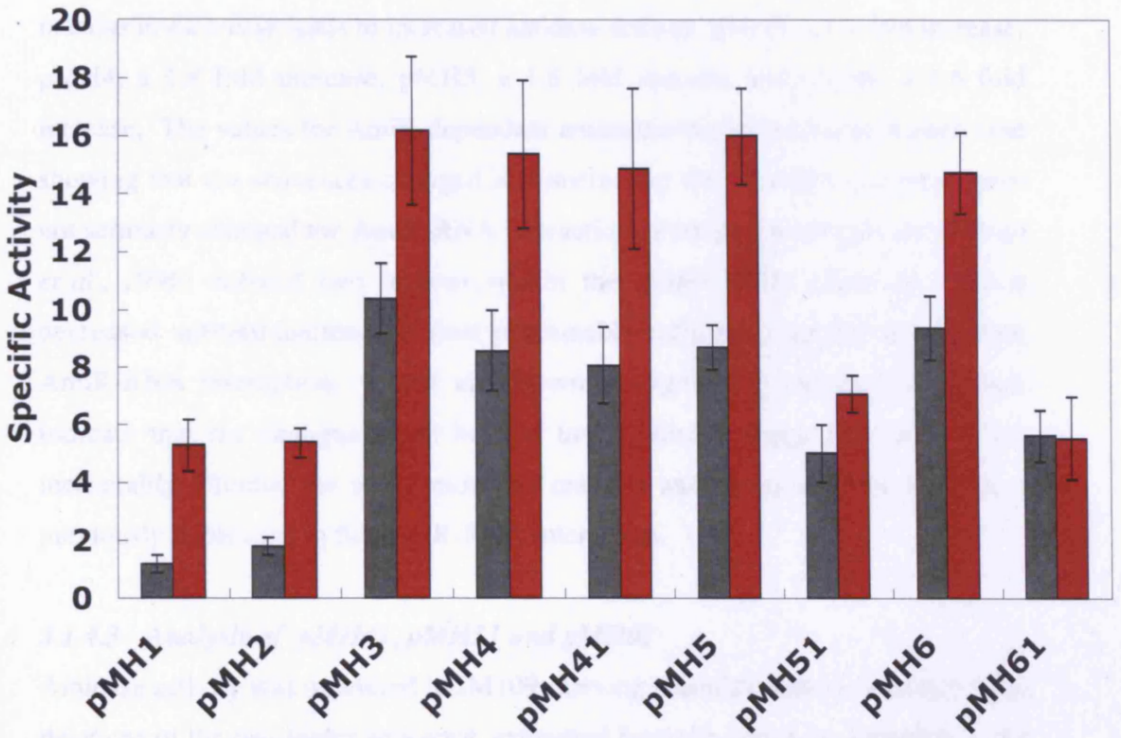


Figure 5.8. Amidase Assays of pGEM4Z Derived Constructs

Assays were carried out in the absence of AmiR [Grey] and presence of AmiR [pSW35] [Red]. Specific activities of pMH1 [1.3 (-amiR), and 5.3 (+amiR)] , pMH2: 1.8 and 5.4, pMH3: 10.4 and 16.3, pMH4 8.6 and 15.4, pM41 8.1 and 14.9, pMH5: 8.7 and 16.0, pMH51 5.0 and 7.1, pMH6 9.5 and 14.7 and pMH61 5.6 (-/+amiR)].

Despite the high background amidase activity from these constructs it is apparent that the leader terminator is still partially functional since in the presence of AmiR *in trans* in each case leads to increased amidase activity: pMH3, a 1.6 fold increase, pMH4, a 1.8 fold increase, pMH5, a 1.8 fold increase and pMH6, a 1.6 fold increase. The values for AmiR-dependent antitermination are similar in each case showing that the sequences changed in constructing the *Hind*III target sites, have not seriously changed the AmiR-RNA interaction. Previous investigations [Wilson *et al.*, 1996] defined two regions within the leader RNA, alterations which decreased antitermination and thus presumably reflected a modified/decreased AmiR-RNA interaction. These are shown in Figure 5.9 and the results here indicate that the changes made here to insert *Hind*III target sites have i] not measurably affected the antitermination reaction and ii] do alter the sequences previously implicated in the AmiR-RNA interaction.

5.1.4.3 Analysis of pMH41, pMH51 and pMH61

Amidase activity was measured in JM109 carrying plasmids with increasingly large deletions of the *ami* leader sequence, expressed from the vector *lac* promoter in the absence and presence of AmiR [pSW35] [Figure 5.8]. The results for pMH41 are comparable to those for the parent plasmid pMH4, giving background activity levels of 8.0 and in the presence of AmiR the levels rise to 15.0. With pMH51, amidase activity is very much reduced in both the absence and presence of AmiR compared to pMH5 but the system still shows some AmiR-dependent antitermination. The deletion in pMH51 has removed almost all *ami* leader sequences before the previously defined L recognition sequence CCGAAC.

Plasmid pMH61, which carries the largest deletion – including the L recognition shows a similar level of background activity to pMH51 – around 5 but no increase in AmiR-dependent antitermination activity. This strongly suggests that some of the RNA sequences essential for AmiR interaction with the leader transcript are missing in this construct such that the antitermination reaction fails.

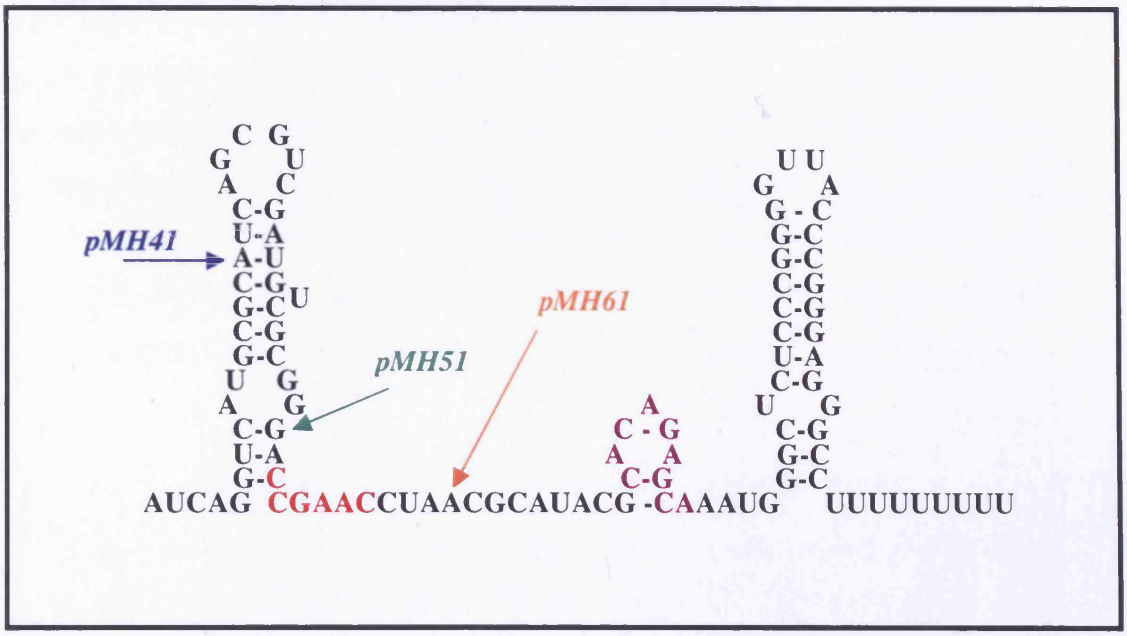


Figure 5.9. RNA fold of the leader region transcript residues 1-100.

The L [Left] and R [Right] recognition sites are shown in red and purple respectively and the start points where the ami leader transcript is fused to the vector lac promoter is shown for plasmids pMH41 [blue], pMH51 [green] and pMH61 [orange].

These results show that removal of *ami* leader transcript sequences up to the L recognition site, does not affect AmiR-dependent antitermination but that removal of the L site prevents antitermination at the down-stream termination loop. These experiments have partially defined a minimal *amiR* transcript, and this knowledge can be used to further investigate the AmiR-RNA interaction.

5.2 *In vitro analysis of the ANTAR -RNA Interaction*

5.2.1 Introduction

Previous experiments, showed the *in vitro* reconstitution of the amide dependent antitermination reaction was achieved by RNA Bandshift using *in vitro* transcribed ^{32}P labelled run-off transcripts of the operon leader region and purified AmiC/AmiR-butyramide complex. The addition of acetamide caused a ligand-dependent bandshift to occur when AmiR bound to the leader RNA. It was found that AmiR binds to the leader RNA with a calculated K_D of 1nM [Norman *et al.*, 2000]. To test the activity of ANTAR *in vitro*, similar experiments were carried out using purified ANTAR and in vitro transcribed ^{32}P labelled run off transcripts.

5.2.2 In vitro Transcription of ami operon Leader Sequence

In vitro transcription reactions are commonly used to produce milligram quantities of RNA from a DNA template [Milligan *et al.*, 1987]. For this study, *in vitro* transcribed ^{32}P labelled leader RNA transcripts were produced to investigate the binding of ANTAR domain to the RNA leader *in vitro*.

Commonly used and best characterised RNA polymerases used *in vitro* transcription reactions are those encoded by the *Salmonella typhimurium* bacteriophage SP6 [Butler and Chamberlain 1982], and the *E.coli* phages T7 and T3 [Studier and Rosenburg 1981; Tabor and Richardson 1985]. The genes for these proteins have been over expressed in *E.coli* and the polymerases have been rigorously purified.

The pGEM4Z plasmid vector used to create the construct pMH41 contains the T7 promoter from which transcription will initiate. Plasmid vectors which are used as templates for *in vitro* transcription are linearized by a restriction enzyme that

creates a 5'-overhang because transcription proceeds to the end of the DNA template, thus linearization prevents the synthesis of multimeric transcripts from a circular DNA template. In this case *Bst*EII was used to linearise the template.

Plasmid pMH41 was considered to contain the shortest length of *ami* leader region which still showed good antitermination and thus presumed AmiR/RNA binding. pMH41 plasmid DNA was digested with *BstE*II, which yielded 3 fragments; 3030, 1011 and 980 bp [Figure 5.10 (a)]. The larger 3030 bp fragment contained the sequence of interest and the T7 promoter region. Because transcription with T7 RNAP will only initiate from the T7 promoter, it was felt that the smaller fragments did not require removal from the DNA template sample. The *ami* leader region contains a single *Bst*EII target located in the centre of the termination loop. Transcription with T7 RNAP from this vector produced an 84-residue run-off transcript [Figure 5.10 (b)].

The *in vitro* transcription reactions were carried out according to a modified version of the MAXIscript® Protocol described in Section [2.2.7.2].

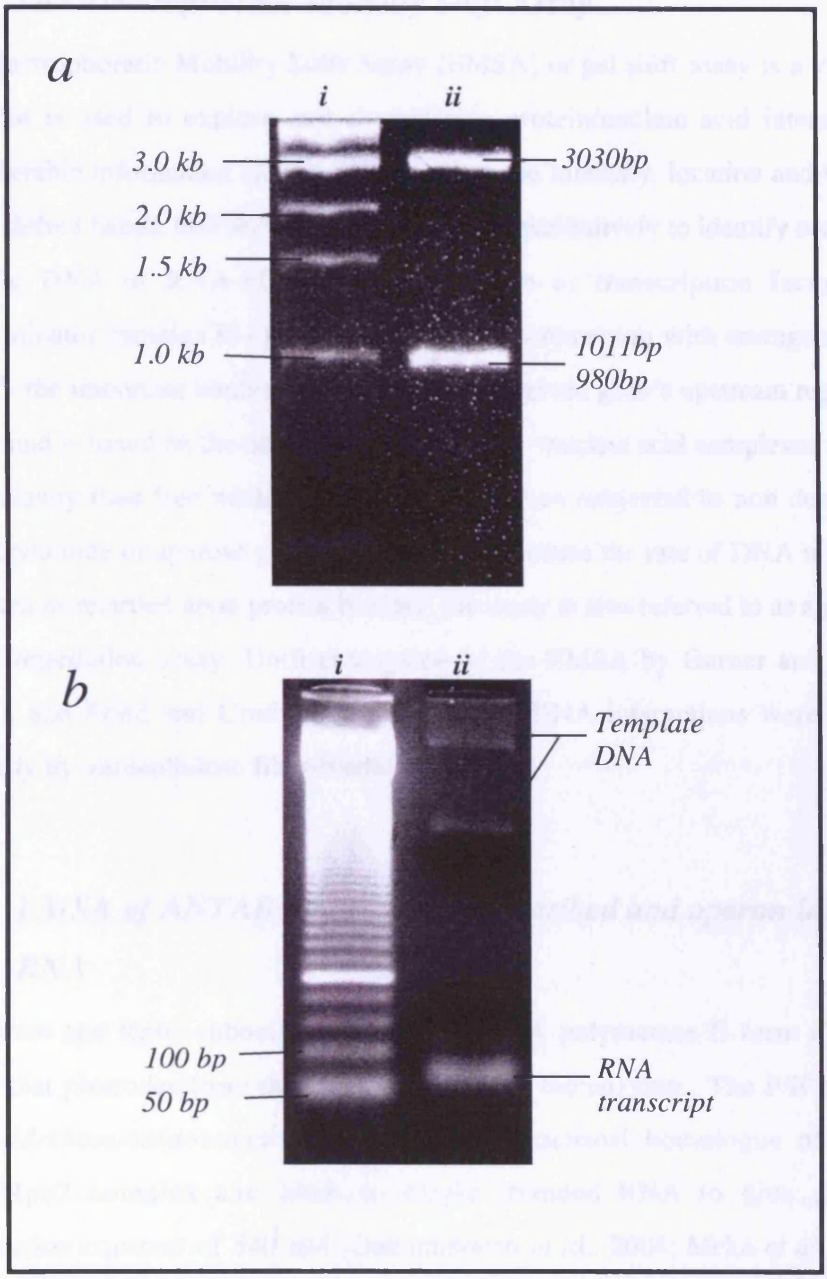


Figure 5.10. Production of <u>in vitro</u> transcribed RNA run off transcripts from pMH41

[a]. BsteII digested pMH41 [<u>amiE</u>/leader] DNA. [b]. The 'Cold' RNA transcript visualised on a 1.2% TBE agarose gel, alongside 50 bp RNA Ladder.

5.2.3 *The Electrophoretic Mobility Shift Assay*

The Electrophoretic Mobility Shift Assay [EMSA] or gel shift assay is a versatile tool that is used to explore and characterize protein/nucleic acid interactions. Considerable information can be obtained from the intensity, location and identity of the shifted bands. Gel shift assays can be used qualitatively to identify sequence-specific DNA or RNA-binding proteins, such as transcription factors and antiterminator proteins in crude lysates and, in conjunction with mutagenesis, to identify the important binding sequences within a given gene's upstream regulatory region and is based on the observation that protein –nucleic acid complexes migrate more slowly than free nucleic acid molecules when subjected to non-denaturing polyacrylamide or agarose gel electrophoresis. Because the rate of DNA migration is shifted or retarded upon protein binding, the assay is also referred to as a gel shift or gel retardation assay. Until conception of the EMSA by Garner and Revzin [1981], and Fried and Crothers [1984], protein:DNA interactions were studied primarily by nitrocellulose filter-binding assays.

5.2.4 *EMSA of ANTAR and in vitro transcribed ami operon leader RNA*

The Rpb4 and Rpb7 subunits of eukaryotic RNA polymerase II form a hetero-dimer that protrudes from the 10-subunit core of the enzyme. The F/E complex from *Methanocaldococcus jannaschii* is a archaeal homologue of human Rpb4/Rpb7 complex and binds to single stranded RNA to give an RNA dissociation constant of 540 nM [Ouhammouch *et al.*, 2004; Meka *et al.*, 2005]. Studies of F/E-RNA complexes using various RNA probes were obtained by Meka and co-workers [2005], with comparable affinities indicating that the binding is not sequence specific and therefore was a suitable protein to use to test the integrity of the *in vitro* transcribed *ami* Leader RNA.

5.2.4.1 EMSA Conditions

For experimental control purposes, archaeal F/E heterodimer and ANTAR were incubated with radiolabelled RNA probe A2, using increasing protein concentrations [0.010 μg, 0.10 μg and 1.0 μg] as described in Section [2.2.7.4]. The ANTAR-RNA samples were then incubated at 37 C and F/E-RNA samples incubated at 65 C for 45 min. The complexes were separated electrophoretically on 4–20% TBE gels [Invitrogen] at 100 V for 2.5 h at 4°C and the images captured using a phosphorimager Figures 5.11 to 5.15

The initial experiment was used to show that the band shifts could occur and the system was functioning correctly. Using Buffer 1 [Table 5.2.] increasing amounts of F/E and ANTAR were incubated with 2500 cpm of A2 RNA probe [Figure 5.11 (a)]. It is apparent from the control lane 1 in which no protein was added that the RNA contained a mixture of fragment lengths. Band shifts are visible in Lanes 3 and 4 containing 0.10 μg and 0.010 μg of F/E. No shifts are observed in the Lanes containing ANTAR.

The same protein concentrations were added to the labelled *ami* leader probe [0.010 μg, 0.10 μg and 1.0 μg] and the products were resolved simultaneously under the same conditions [Figure 5.11 (b)]. This probe in the absence of protein [Lane 1] ran as a number of discrete bands possibly due to secondary structure formation. As with the A2 RNA probe, the *ami* leader probe band shifts were only visible with the highest F/E protein concentration [1.0 μg] and no shifts were observed with ANTAR.

The A2 RNA probe and F/E protein was a gift from Finn Werner at UCL .

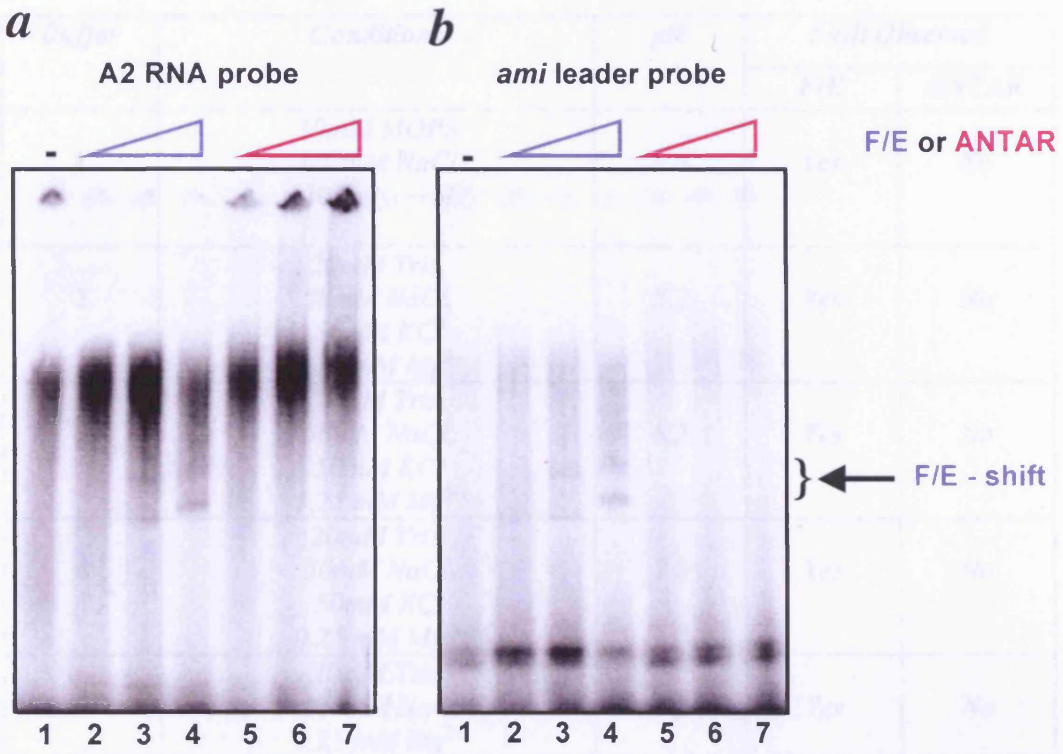


Figure 5.11. RNA EMSA using Buffer Conditions 1

[a] EMSA using the A2 RNA probe [2500 cpm] and F/E protein [lanes indicated by blue triangles] of M. jannashii. Lane 1 control no additions, Lane 2, 0.010 μg F/E, Lane 3, 0.10 μg F/E, Lane 4, 1.0 μg F/E. Lanes 5 to 7 [indicated by pink triangles] contain A2 probe with ANTAR concentrations of 0.010 μg, 0.10 μg and 1.0 μg respectively. [b] EMSA using the ami leader RNA probe [2500 cpm] and F/E protein [lanes indicated by blue triangles]. Lane 1, control no additions, Lane 2, 0.010 μg F/E, Lane 3, 0.10 μg F/E, Lane 4, 1.0 μg F/E. Lanes 5 to 7 [indicated by pink triangles] contain ami leader probe with ANTAR concentrations of 0.010 μg, 0.10 μg and 1.0 μg respectively.

Table 5.2. Buffer conditions used for the EMSA of ANTAR and ami Leader

Buffer	*Conditions*	*pH*	*Shift Observed*	
			F/E	*ANTAR*
1	*10mM MOPS* *100mM NaCl* *10%glycerol*	*6.8*	*Yes*	*No*
2	*20mM Tris,* *50mM NaCl,* *50mM KCl* *0.25 mM Mg^{2+}*	*8.2*	*Yes*	*No*
3	*20mM Tris,* *50mM NaCl,* *50mM KCl* *1.25 mM Mg^{2+}*	*8.2*	*Yes*	*No*
4	*20mM Tris,* *50mM NaCl,* *50mM KCl* *0.25 mM Mg^{2+}*	*7.2*	*Yes*	*No*
5	*20mM Tris.* *100mM NaCl* *0.25 mM Mg^{2+}*	*9.2*	*Yes*	*No*

The initial experiments showed that a discreet labelled *ami* leader transcript had been produced and the reaction conditions were such that a non-specific RNA binding protein F/E could cause RNA band shifting. Under these conditions, no ANTAR binding was detectable. The experiment were repeated using only the *ami* leader RNA with both F/E and ANTAR proteins under four additional different buffer conditions to try and establish specific ANTAR/RNA binding.

The reactions using Buffer 2 [Table 5.2] are shown In Figure 5.12. These are the conditions under which AmiR-dependent bandshifting was shown previously using the AmiC/AmiR-butyramide complex. The complex was dissociated using acetamide in the presence of RNA and bandshifts were observed [Norman et al.,

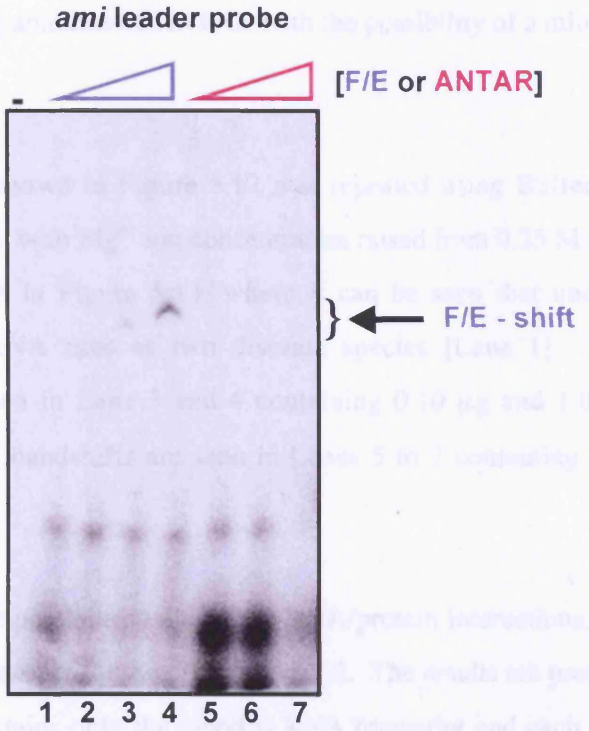


Figure 5.12. RNA EMSA using Buffer Conditions 2

EMSA using the ami leader RNA probe [2500 cpm] and F/E protein [lanes indicated by blue triangles]. Lane 1 control no additions, Lane 2, 0.010 μg F/E, Lane 3, 0.10 μg F/E, Lane 4, 1.0 μg F/E. Lanes 5 to 7 [indicated by pink triangles] contain ami leader probe with ANTAR concentrations of 0.010 μg, 0.10 μg and 1.0 μg respectively.

2000]. As previously the *ami* leader RNA ran as a number of discrete species in the absence of protein [Lane 1]. In the presence of F/E band shifts were visible in Lane 3 and 4 containing 0.10 μg and 1.0 μg of protein respectively. Lanes 5 to 7 contain increasing amounts of ANTAR with the possibility of a minor shift in Lane 7.

The experiment shown in Figure 5.12 was repeated using Buffer 3, based on a modified Buffer 2, with Mg^{2+} ion concentration raised from 0.25 M to 1.25 M. The results are shown in Figure 5.13, where it can be seen that under these assay conditions the RNA runs as two discrete species [Lane 1]. F/E dependent bandshifts are seen in Lane 3 and 4 containing 0.10 μg and 1.0 μg of protein respectively. No bandshifts are seen in Lanes 5 to 7 containing added ANTAR protein.

To investigate the pH dependence of the RNA/protein interactions, Buffer 4 is the same as Buffer 2 except for the pH value of 7.2. The results are presented in Figure 5.14. Lane 1 contains only the labelled RNA transcript and each lane containing F/E shows some degree of bandshifting. There is no evidence of a discrete ANTAR/leader RNA interaction at any concentration used [Lanes 5 to 7].

The final Buffer used [Buffer 5] was at pH 9.2 and contained only Tris [20mM] NaCl [100mM] and Mg2+ ions [0.25mM] and the results can be seen in Figure 5.15. As previously, bandshifting was seen with the F/E protein but not with ANTAR.

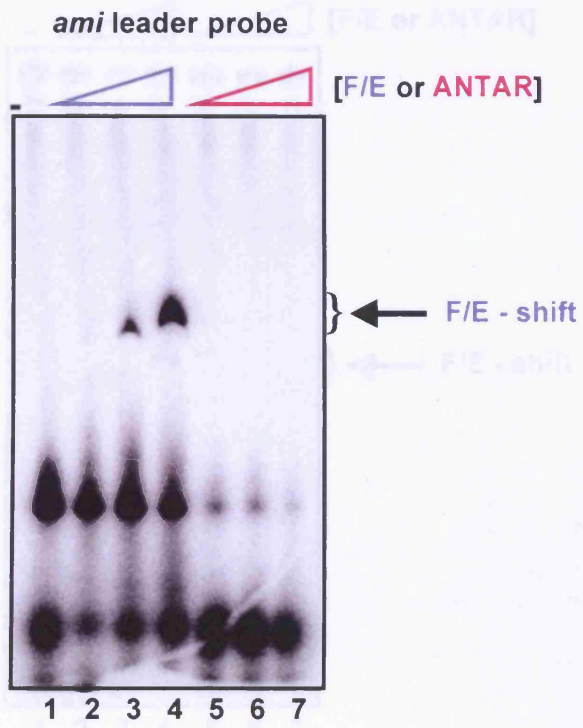


Figure 5.13. RNA EMSA with buffer Conditions 3

EMSA using the ami leader RNA probe [2500 cpm] and F/E protein [lanes indicated by grey triangles]. Lane 1, control no additions, Lane 2, 0.010 µg F/E, Lane 3, 0.10 µg F/E, Lane 4, 1.0 µg F/E. Lanes 5 to 7 [indicated by pink triangles] contain ami leader probe with ANTAR concentrations of 0.010 µg, 0.10 µg and 1.0 µg respectively.

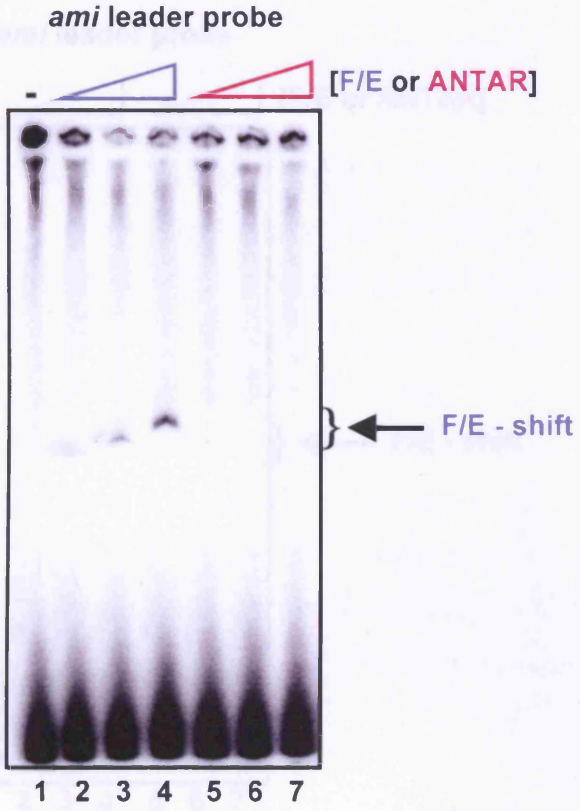


Figure 5.14. RNA EMSA using buffer Conditions 4

EMSA using the ami leader RNA probe [2500 cpm] and F/E protein [lanes indicated by blue triangles]. Lane 1 control no additions, Lane 2, 0.010 μg F/E, Lane 3, 0.10 μg F/E, Lane 4, 1.0 μg F/E. Lanes 5 to 7 [indicated by pink triangles] contain ami leader probe with ANTAR concentrations of 0.010 μg, 0.10 μg and 1.0 μg respectively.

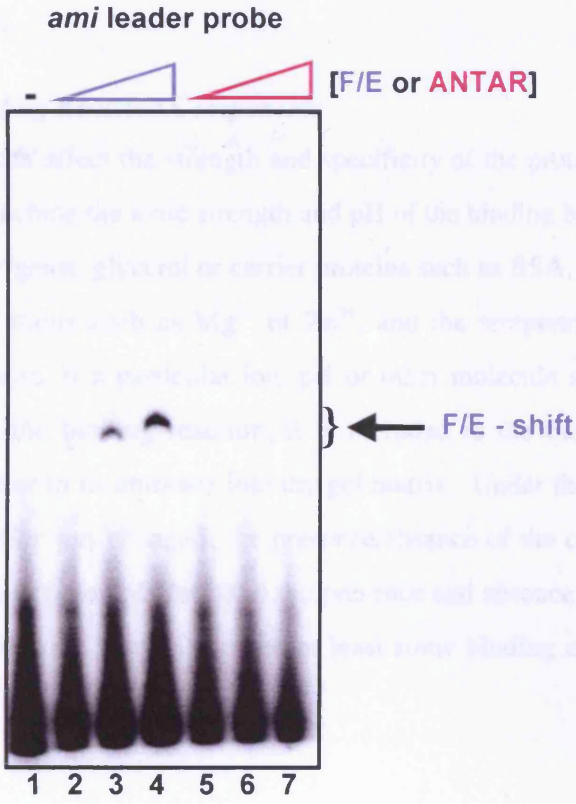


Figure 5.15. RNA EMSA using Buffer Conditions 5

EMSA using the ami leader RNA probe [2500 cpm] and F/E protein [lanes indicated by blue triangles]. Lane 1 control no additions, Lane 2, 0.010 µg F/E, Lane 3, 0.10 µg F/E, Lane 4, 1.0 µg F/E. Lanes 5 to 7 [indicated by pink triangles] contain ami leader probe with ANTAR concentrations of 0.010 µg, 0.10 µg and 1.0 µg respectively.

5.2.5 ANTAR does not exhibit RNA Binding Activity under these conditions.

The results from the RNA-EMSAs show that the ANTAR domain of AmiR does not bind to RNA when tested under these buffer conditions in which F/E was shown to bind.

5.2.5.1 Binding Reaction Components

Factors that can affect the strength and specificity of the protein:RNA interactions under study include the ionic strength and pH of the binding buffer, the presence of nonionic detergents, glycerol or carrier proteins such as BSA, the presence/absence of divalent cations such as Mg^{2+} or Zn^{2+}, and the temperature and time of the binding reaction. If a particular ion, pH or other molecule is critical to complex formation in the binding reaction, it is included in the buffers to stabilize the interaction prior to its entrance into the gel matrix. Under the conditions used the pH of the buffer was changed, the presence/absence of the concentration of Mg^{2+} ions, the concentration of NaCl and the presence and absence of KCl. Under these conditions used, F/E always showed at least some binding and ANTAR failed to bind.

5.3 *In vivo analysis of the ANTAR -RNA Interaction*

5.3.1 *Introduction*

Despite a failure to detect ANTAR-RNA binding activity *in vitro*, it was deemed necessary to additionally look for antitermination activity *in vivo*. For the purposes of an *in vivo* assay, ANTAR was cloned into pETDuet-1, with the aim of co-expressing this domain with the *amiE*/Leader sequence expressed from the same plasmid [Table 5.3]. As described in 3.2.3.1 the vector contains two MCSs, both of which are preceded by a T7 promoter/*lac* operator and RBS. The advantages of dual expression from a single plasmid has been previously discussed and this pET System was used once again to test levels of antitermination by ANTAR by carrying out amidase assays as a measure of antitermination by ANTAR. It is also possible to use pETDuet in conjunction with pACYCDuet, enabling the coexpression of two further proteins if required. Both pETDuet and pACYCDuet are designed with compatible replicons [*ColE1, P15A*] and drug resistance genes [*bla, cat*] for the effective propagation and maintenance of two plasmids in a single cell. This opens up the possiblilty of testing other domains within AmiR which may work in conjunction with ANTAR to cause antitermination.

Table 5.3. pETDuet-1 derivatives Constructed for this chapter.

Plasmid	Cloning Vector	Ab	Phenotype
pMH902	pETDuet-1	Amp	*Ami* leader +*amiE* gene MCS1, *antar* MCS2
pMH92	pETDuet-1	Amp	*antar* MCS2

5.3.2 Cloning of ANTAR into pETDuet-1 and pMH90

The *amiE* gene/*ami*Leader sequence had previously been cloned into pETDuet MCS2 creating pMH90 and *antar* was cloned into MCS1 of pMH90 to create pMH902. A control plasmid expressing only ANTAR was created by cloning *antar* into the MCS1 of pETDuet-1, to create pMH92 [Figure 5.16]. For cloning of *antar* into pETDuet-1, primers with incorporated *Nde*I [Forward primer] and *Xho*I [Reverse primer] sequences were used to amplify *antar* in a PCR reaction from pSW24. The subsequent PCR product was digested with *Nde*I and *Xho*I for 1 hour at 37 C then heated to 65 C to denature the enzymes before ligation into *Nde*I and *Xho*I cut pETDuet-1 and pMH90 to create pMH92 and pMH902 respectively.

Table 5.4. PCR primers used for the cloning of antar into pETDuet-1

Name	*Sequence [5' –3']*	*Features*
AntarFor1	TATA**CATATG**CGGATC AACCAGGCCAAG	*Nde*I [CATATG] **Initiator ATG**
AntarRev1	TTAA**CTCGAG**TCAGGCG GACGGCTCGTTTCCCAG	***Xho*I [CTCGAG]**

Plasmids pMH92 [*antar*] and pMH902 [*amiE, antar*], were transformed into BL21 DE3, pLysS and assayed for amidase activity under non-inducing and inducing [+IPTG] conditions alongside the strain carrying pMH90 [*amiE*], pMH91 [*amiR*] and pMH901[*amiE, amiR*] [Figure 5.17]. As previously shown, [3.2.3.2], both pMH90 and pMH91 show background colour production due to non-specific esterase activity and pMH901 shows high background amidase activity [SA 4.0] in the absence of IPTG and SA of 13.2 upon T7 RNA polymerase production. Both pMH92 and pMH902 show only background esterase activity showing that the ANTAR domain has no *in vivo* antitermination activity.

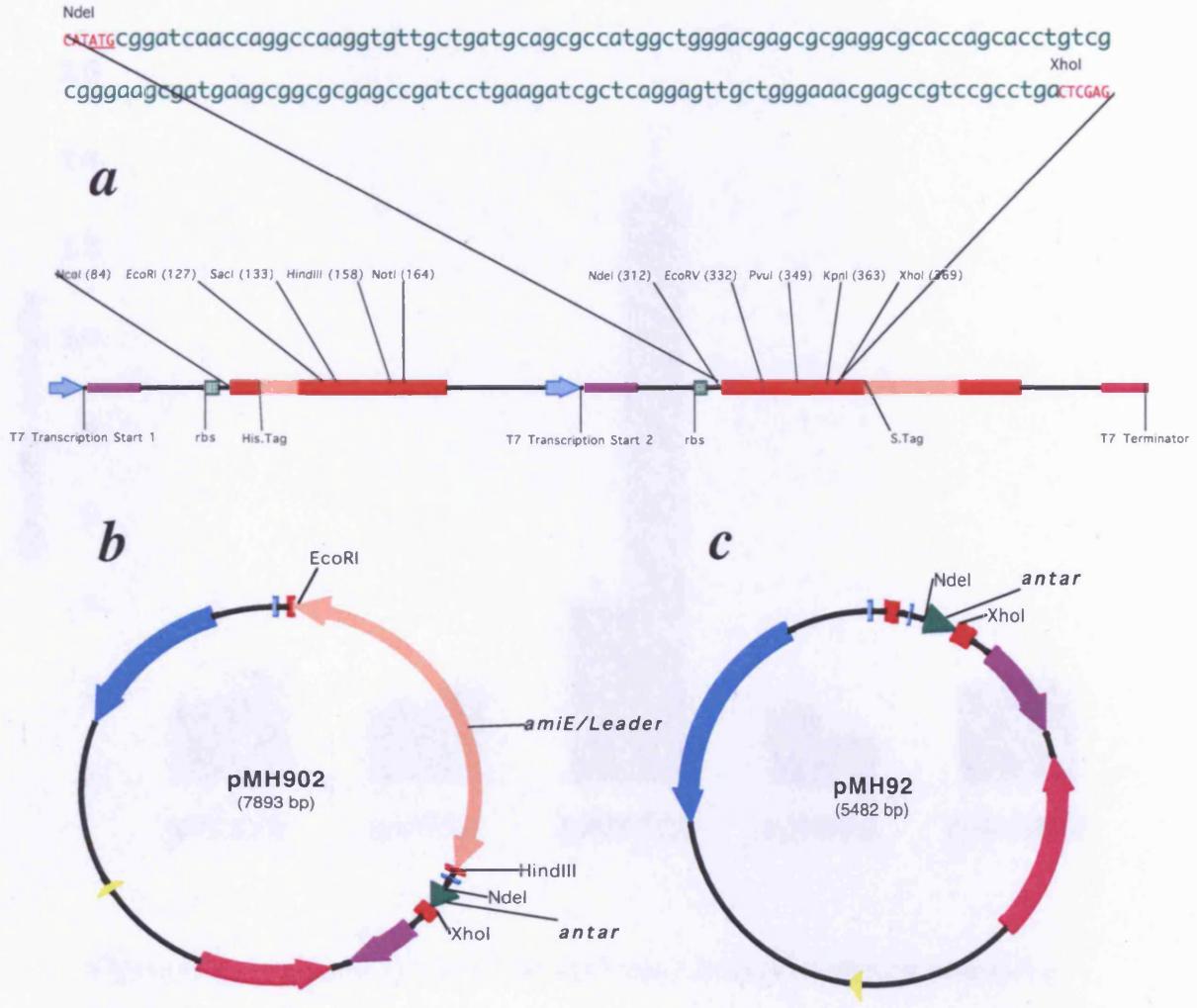


Figure 5.16. Construction of pMH902 and pMH92 derivatives of pETDuet-1.

[a] Part of the pETDuet sequence showing both MCSs and the single stranded ANTAR sequence cloned into MCS1 to create pMH92. [b] Structural map of pMH902, which has the amiE leader sequence cloned into MCS1 and the antar sequence [green] cloned into NdeI XhoI RE sites within MCS2 and [c] Structural map of pMH92, which has the antar sequence in MCS1 only.

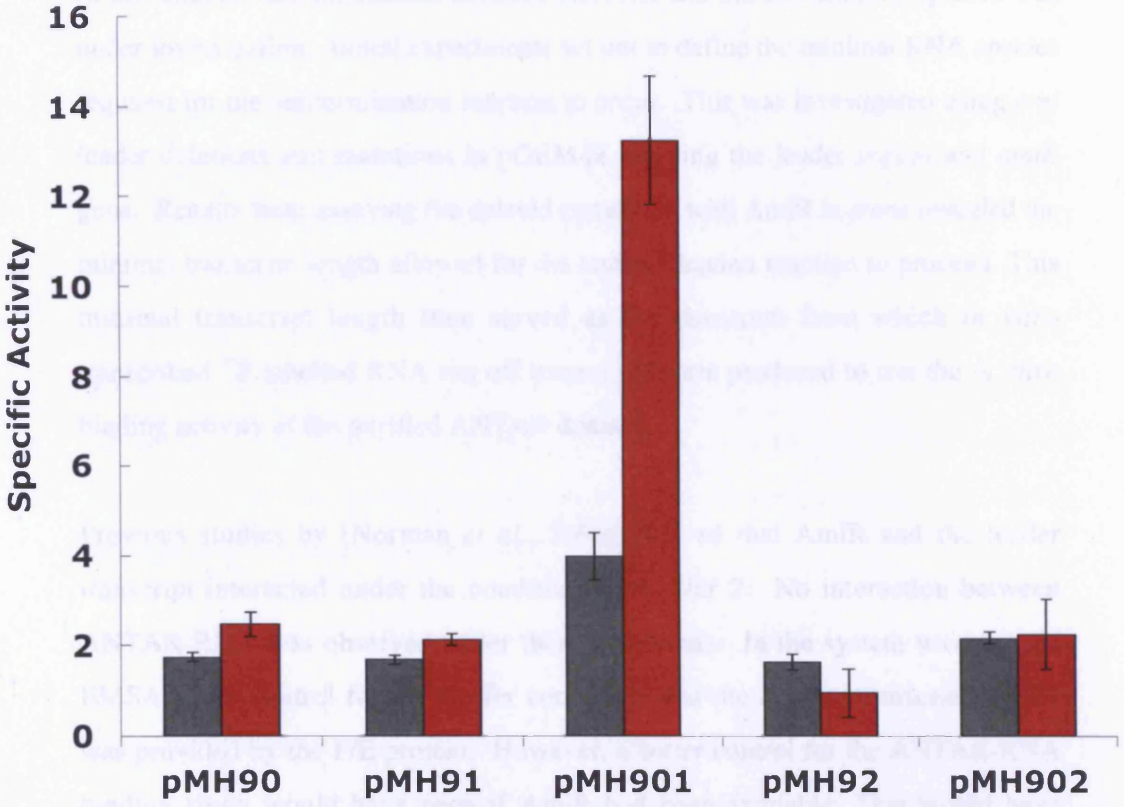


Figure 5.17. Amidase assays of the pETDuet-1 derived constructs containing AmiR and ANTAR

Assays were carried of plasmids in Bl21 DE3 pLysS E. coli cells. Specific activities of uninduced [grey (-)] and 1mM IPTG induced [red (+)] pMH90 [SA 1.7(-), 2.5 (+)], pMH91 [SA 1.6 (-), 2.1 (+)], pMH901 [SA 4.0 (-), 13.2 (+)], pMH92 [SA 1.6 (-), 0.9 (+)], and pMH902 [SA 2.2 (-), 2.3 (+)].

5.3.3 Summary of Chapter 5

In this chapter, the interaction between ANTAR and the *ami* leader sequence was under investigation. Initial experiments set out to define the minimal RNA species required for the antitermination reaction to occur. This was investigated using *ami* leader deletions and mutations in pGEM4Z carrying the leader region and *amiE* gene. Results from assaying the deleted constructs with AmiR *in trans* revealed the minimal transcript length allowed for the antitermination reaction to proceed. This minimal transcript length then served as the construct from which *in vitro* transcribed ^{32}P labelled RNA run off transcripts were produced to test the *in vitro* binding activity of the purified ANTAR domain.

Previous studies by [Norman *et al.*, 2000] showed that AmiR and the leader transcript interacted under the conditions in Buffer 2. No interaction between ANTAR-RNA was observed under these conditions. In the system used for the EMSA's, the control for the Buffer conditions and the *in vitro* transcribed RNA was provided by the F/E protein. However, a better control for the ANTAR-RNA binding study would have been if AmiR had been available. This would have shown that AmiR does indeed bind/or does not bind under those buffer conditions and would have provided a system were a direct comparison could be made between the results.

6 Discussion

Before embarking upon this research project, much was known about the regulatory genes AmiC and AmiR, and the method by which the pair controlled amidase operon expression. The regulation mechanism adopted by AmiC and AmiR was fully elucidated when the crystal structures of AmiC-acetamide [Pearl *et al.*, 1994] and the AmiC-AmiR complex [O' Hara *et al.*, 1999] were solved. AmiC, the negative controller of amidase operon expression exerted its control of AmiR via a steric hindrance mechanism. Despite this model, little was known about how AmiR antiterminates transcription. The obvious next step was to solve the crystal structure of the AmiR/RNA complex.

For many years and despite repeated attempts to purify AmiR, the protein remained insoluble and impossible to purify [R.Drew personal communication]. The overall structure of AmiR is unusual due to the highly hydrophobic, long 50-residue α-helix, which separates the globular N-terminal domain from the C-terminus. In the AmiC/AmiR complex, AmiR exists as a dimer and the extremely strong hydrophobic interactions which cause dimerisation suggest it is unlikely to exist in a monomeric form. The N-terminal domain consists of approximately 100 amino acids and is folded into an alternating α/β structure with an open twisted parallel β–sheet in the middle surrounded by α–helices on both sides. This type of structure is known as a Rossman fold [Rao and Rossman, 1973]. In the AmiC/AmiR complex, the Rossman fold is occluded, suggesting that the RNA binding activity could reside here. The AmiR C-terminus contains a helix-turn-helix motif known as an ANTAR domain [Shu and Zhulin, 2002]. Previous work carried out by O'Hara and others [1999], showed that deletions to the last two helices of the CTD of AmiR resulted in an uninducible phenotype – implying the

antiterminatory activity resided here. Sequence analysis carried out by Shu and Zhulin [2002] and Morth and co-workers [2004] identified four highly conserved alanine residues [Ala 152, Ala 167, Ala 175, and Ala 186], and one aromatic residue, which in ANTAR is Histidine.

Due to the difficulties presented by the insolubility of AmiR, the putative RNA-binding ANTAR domain was the subject of investigation in this thesis.

6.1 Mutagenesis of AmiR

Site Directed Mutagenesis was carried out to change three of these conserved residues and amidase assay results of these mutations showed that A167 and H168 are important for the antitermination ability of AmiR. A152 mutated to V and G, results in AmiR having reduced activity, 22.0 and 18.6 respectively compared to WT of 37.0. Therefore, although antitermination is reduced, AmiR is still able to antiterminate.

Mapping of these residues shows that the side chains of the Alanine residues all point inward to the cavity formed by overall architecture of the ANTAR domain and it is likely that their conservation is important in maintaining this tertiary structure. The aromatic group of the H168 residue points upwards and the possibility exists that it is involved with interacting with RNA during the antitermination reaction. Histidine residues have been in the past shown to play an important role in protein-RNA interactions, contributing to stacking interactions with nucleic acid bases [Baker and Grant, 2007]. The close proximity of A167 to the essential H168 may contribute to the functioning of H168, and a mutation to A167 could effect the steric interactions it has.

For a more meaningful mutational analysis of the conserved residues within ANTAR, double mutational analysis would have been useful in determining if the single amino acid changes, which have been shown to affect antitermination, have a synergistic effect when produced in combinations. For example producing a double mutant; A167G + H168A and comparing the results from amidase assays of the single A167G and H168A mutants would tell us if the loss in antitermination activity is greater in combination than the sum of its parts; indirectly indicating that the two mutated residues work in combination rather than having separate affects in terms of binding to the ligand. This could have been achieved by using SDM and the H168A mutant plasmid as the parent for further mutations. This form of mutational analysis was not carried out due to a lack of time and the results from the single point mutational analysis thus far were deemed sufficient enough to justify purifying the ANTAR domain and pursuing the purification of the ANTAR/RNA complex.

Random mutagenesis of *amiR* was used to discover if there are other areas of the protein required for the antitermination reaction, and to detect other important residues within ANTAR by selecting for and analyzing antitermination defective mutants. A number of strategies were tried for the random screen and the system finally chosen was the use of sloppy PCR and screening using a pETDuet derivative. In this system the mutated *amiR* and *amiE*/Leader were expressed from a single plasmid. This single plasmid system, developed in this study for the mutagenesis screen, was a novel approach in AmiR+*ami*/Leader interaction studies and proved to be a very reliable system in which to test such intermolecular interactions, its application in future experiments should be advantageous. The construct, pMH901 showed that the production of AmiR leads to antitermination and subsequent production of amidase. A control plasmid bearing only the *amiR* gene in pETDuet-1 [pMH91] exhibited no amidase activity. To test that pMH91 was indeed producing functional – antiterminatable AmiR, it was assayed in the

presence of *amiE*/Leader on a second plasmid, pTM1, which gave an amidase SA value of 11.2. A second control plasmid bearing only *amiE*/Leader [pMH90] showed that in the absence of AmiR, the majority of transcripts terminated, thus the amidase activity exhibited by pMH901 was due to antitermination by AmiR interacting with the leader sequence.

A drawback of this system was that due to lack of funding, only a limited number of clones identified in the screen could be sequenced, therefore it was difficult to fully optimise the method. Despite this, the random mutagenesis studies of AmiR carried out in this chapter led to the discovery of residues throughout the protein including ANTAR, which are required for AmiR to fully function as an antiterminator protein. The V154E mutation affects a residue buried between the two AmiR molecules towards the bottom of the long coiled coil [ANTAR] and contacts the M157 and E164 on the other dimer member. Altering the hydropathicity profile within this region may lead to instability of the dimer. The R179H mutation lies within a group of charged residues [KRRE] making the turn from helix 7 to helix 8. Alteration to this residue may affect the turn from the one helix to the next. The P20 residue in the NTD of AmiR is found with the sequence LNPPG in the loop between the end of a β-strand and beginning of helix 2. The P20Q change identified may cause a structural change to the NTD. This data has provided further support to the theory that ANTAR has RNA binding activity.

If further work using random mutagenesis system could have been carried out, it would be of value to further investigate the mutants uncovered in this study by identifying restorative residues in a given antitermination defective protein sequence. Mutations that suppress or compensate the defects introduced in this assay could be analysed. Using an antitermination defective mutant as the parent plasmid in the mutagenic PCR reaction and then selecting for strains, which appear to restore the antitermination levels to that of WT, could possibly tell us if the two

residues work in combination and provide insight into the nature of the structural basis for the effects of the initial mutation.

The above interpretations of the reduced antitermination abilities of the mutated AmiR proteins depend on the levels of expression of these mutated proteins being the same as those of wild-type AmiR. It is possible that the introduced mutations reduced expression levels by destabilizing AmiR or through indirect effects on transcription or translation.

Quantitation of the expressed mutant AmiR proteins, by Western Blotting using AmiR specific antibodies [Norman *et al.*, 2000], or by scanning stained SDS-PAGE gels, would confirm mutant AmiR proteins are being expressed at the same level as the wild-type AmiR and instability with accompanying proteolytic degradation may be indicated by observing additional bands on Western Blots. Codon usage uncommon to the expressing bacteria can also reduce expression levels. This can be ruled out by making individual expression plasmids for all possible codon combinations.

The possibility also exists that an antitermination deficient mutant AmiR may enhance the termination reaction when interacting with the leader mRNA, thereby preventing expression, rather than just lacking antitermination activity. Both would result in reduced specific activities. Due to time constraints, a method was not developed to differentiate between the two.

6.2 Characterisation of ANTAR

For the first time the AmiR ANTAR domain has been cloned and expressed in vector pGEX6P1 and purified as a cleavable, GST-tagged protein. Hydropathicity plots had predicted that ANTAR is hydrophilic compared to the rest

of the protein, and as predicted, purification of this entity proved to be easier than AmiR had reported to be. Some contaminants had co-eluted during the purification of the GST tagged protein, thus, one further purification step was required to obtain fairly pure sample and ANTAR was further purified using gel filtration chromatography. Data from the gel filtration elution profile of ANTAR show that it elutes as a dimer, and this was confirmed using DLS. DLS data gives the average molecular weight for ANTAR as 10.5 kDa. The calculated MW of ANTAR is 5.5 kDa. The DLS percentage polydispersity calculation indicates that the ANTAR protein is mono-dispersed in solution and is not forming aggregations.

The activity of ANTAR as an RNA binding protein was investigated *in vivo* and *in vitro* and the minimal RNA species required for the antitermination reaction to occur was established. RNA-binding experiments were carried out in the form of EMSAs. The method was established using the F/E protein and A2 RNA probe, which are known to produce a bandshift under Buffer 1 conditions [Meka *et al.*, 2002] as a control. ANTAR was incubated with ^{32}P labelled *in vitro* transcribed RNA directed from pMH41, which had previously been shown in chapter 5.1 to produce the minimal transcripts for AmiR antitermination. Buffer conditions 2 were based on those used previously for AmiR leader titration [Norman *et al.*, 2000]. No ANTAR-dependent shift was seen under these conditions. The conditions were then adjusted to suit the ANTAR protein - and the calculated pI of ANTAR is 8.7, therefore the pH of Buffer 2 and Buffer 3 was altered to 9.2 and 7.2 respectively. In addition to this, the Mg^{2+} conditions were increased from 0.25 mM to 1.25 mM. In none of these Buffer conditions was a shift demonstrated with the ANTAR-RNA.

The EMSA assay is used for the detection of RNA and protein complexes, and is used more favourably when there is strong binding between the two. This is an important feature of EMSA; once the complex forms, it has to remain sufficiently complexed when migrating through the gel during electrophoresis. In these

experiments, formation of a complex between ANTAR and RNA cannot be strictly ruled out because the possibility exists that the complex disintegrates during electrophoresis, - therefore not causing a shift.

Prior to utilising the EMSA assay, an alternative RNA binding method was used to investigate the ANTAR-RNA interaction. A fluorometric reporter approach was used, [the results of which have not reported in this thesis due to efforts being abandoned in favour of developing the EMSA assay]. For this experiment, the entire 100 bp *ami* leader RNA sequence was purchased [Dharmacon] with a biotin label at the 5' end. The ANTAR protein was used to coat the required number of wells of black [for low fluorescence], high protein binding capacity, 96-well plates [Biomat], followed by blocking with BSA and washing steps. These plates were then incubated with the biotinylated leader RNA, followed by washing to remove unbound biotinylated RNA. The ANTAR biotinylated-RNA complex formation was measured fluorometrically using a FluoReporter Biotin Quantitation Assay Kit on a fluorescence microplate reader [FLUOstar OPTIMA]. This kit contains a fluorogenic Biotective Green reagent which exhibits bright green fluorescence in the presence of biotin. Biotin was not detected in any of the wells containing ANTAR. After the first two attempts this method was abandoned in favour of the EMSA method because it required smaller amounts of protein and the fluorometric method offered no advantage; both strategies are reliant on strong binding between protein and RNA and are not good for detecting complexes based on weak interactions.

If the complex between RNA and ANTAR is weakly associated, then a method not reliant on the ability to preserve stable interactions could detect formation of a complex between these two species. Analytical Ultracentrifugation and Circular Dichroism spectroscopy [CD Spectroscopy] are two such techniques.

Sedimentation velocity analysis using an analytical ultracentrifuge can be used to detect complex formation by comparisons of the sedimentation coefficients of free ANTAR and RNA bound ANTAR. Complex formation between ANTAR and RNA would result in an increase in mass and this would be reflected in the sedimentation behaviour of ANTAR. CD spectroscopy could also have been used for the detecting interaction between ANTAR and RNA; structural changes in ANTAR, upon RNA binding would be reflected in changes in the CD spectra.

The results from the *in vivo* analysis of the ANTAR antitermination reaction showed that ANTAR does not have any *in vivo* activity. However, it was hoped that ANTAR would still bind to the RNA and it would have been possible to purify a stable complex for further structural analysis.

Before drawing assumptions from these data, it is important to consider that the binding conditions of ANTAR may be dramatically different to that of the AmiR-RNA interaction. In addition, the purified ANTAR species has the additional amino acid residues GPLSG on the N-terminus, and this may hamper the binding that would normally occur. The 49-residue, 5.5 kDa ANTAR domain is a relatively small species compared to AmiR, representing 25% of the AmiR protein. The added GPLSG may have had a detrimental effect on the RNA interaction. However, despite the residual GST tag, expression from pGEX6P1, which offers cleavage by Precission Protease appeared to be the best option, as this protease left the smallest number of residual amino acids post cleavage.

The possibility also exists that the ANTAR species purified does not represent the form in which ANTAR is able to bind. The AmiR molecule possibly undergoes a structural rearrangement or oligomerisation prior to RNA binding following its release from AmiC. Perhaps to fully rearrange into its active form other parts of the protein are required. Recall the structure of RV1626 [Morth *et al.*, 2004],

which shows a high level of structural similarity with AmiR and same overall fold of the NTD and CT ANTAR domain. The long hydrophobic alpha helix folds in on itself, rather than forming a dimer with another RV1626 molecule as occurs with AmiR in the AmiC-AmiR complex. This folded-over conformation could reflect the active conformation of AmiR and would not be possible with just the ANTAR domain.

The results presented in this thesis have shown that the ANTAR domain is required for the antitermination reaction to occur. The mutagenesis experiments in this study give weight to the theory that ANTAR is essential in the antitermination reaction. Residues A152, A167 and H168 all are important in the reaction. However, ANTAR alone is not sufficient to bind RNA or cause antitermination.

6.3 Possibilities For Future Experimentation

6.3.1 Purifying the AmiR/RNA Complex

The detailed visualisation of the intact machinery of a protein-RNA complex presents a major challenge in structural biology. The numbers of protein-RNA structures that have been solved are rapidly increasing, although not at as fast a rate as protein-DNA structures. The major obstacle appears to be the delicate nature of the RNA molecule and difficulties that arise when crystallising RNA molecules. In the case of AmiR, the insolubility and hydrophobic nature of the protein makes this task much more difficult.

In addition to the further work suggested above; the double mutational analysis, restorative random mutagenesis studies, and the weak binding assays with ANTAR and *in vitro* transcribed RNA, purification of AmiR/RNA complex is the best and

most obvious avenue by which to investigate the mechanism AmiR antiterminates transcription. There are various methods by which this may be achieved.

6.3.1.1 Purification of AmiR-RNA complex formed in vivo

There are several vectors available that allow the expression of two genes simultaneously, under the control of one plasmid. In Chapter 3, this system was used to express high levels of fully functional AmiR, which interacted with the *ami* leader sequence and caused antitermination. A system similar to this could be employed to express large amounts of AmiR and equally large amounts of *ami* leader, which would titrate out AmiR. Together, the complex may be more soluble if AmiR is more stable when using the RNA as a scaffold. Furthermore AmiR could be tagged in this system allowing the possibility of easier purification of the complex.

6.3.1.2 Purification of AmiR-RNA complex formed in vitro

Using *in vitro* transcribed *ami* leader RNA molecules, the AmiR complex could be purified with the RNA bound to a column and used to capture AmiR during protein purification.

6.4 Conclusion

One of the fundamental challenges of molecular biology is to understand the transcriptional regulation of gene expression. We are a step closer in the elucidation of the antitermination mechanism by AmiR in the amidase operon. The field of protein purification is ever evolving and improving and the way forward is to put effort into finding a method to purify AmiR in complex with its RNA ligand.

7 References

Adelman, J.L., Jeong, Y.J., Liao, J.C., Patel, G., Kim, D.E., Oster, G., and Patel, S.S. [2006]. Mechanochemistry of Transcription Termination Factor *Rho*. Mol. Cell. *22*, 611-621.

Adhya, S., Gottesman, M., and De Crombrugghe, B. [1974]. Release of polarity in *Escherichia coli* by gene N of phage lambda: termination and antitermination of transcription. Proc. Natl. Acad. Sci. U S A. *71*, 2534-2538.

Allison, T.J., Wood, T.C., Briercheck, D.M., Rastinejad, F., Richardson, J.P., and Rule G.S. [1998]. Crystal structure of the RNA-binding domain from transcription termination factor *rho*. Nat.Struct.Biol. *5*, 352-356.

Amster-Choder, O., and Wright, A. [1993]. Transcriptional Regulation of the *bgl* operon of *Escherichia coli* involves phosphotransferase system-mediated phosphorylation of a transcriptional antiterminator. J. Cell. Biochem. *51*, 83-90.

Arnvig, K.B., Pennell, S., Gopal, B., and Colston, M.J. [2004]. A high-affinity interaction between NusA and the rrn nut site in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U S A. *101*, 8325-8330.

Aymerich, and Steinmetz. [1992]. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. Proc. Natl. Acad. Sci. U S A. *89*, 10410–10414.

Baker, C.M. and Grant, G.H. [2007]. Role of aromatic amino acids in protein-

nucleic acid recognition. Biopolymers. *85*, 456-470.

Barnard, A., Wolfe, A., and Busby, S. [2004]. Regulation at complex bacterial promoters: how bacteria use different promoter organisations to produce different regulatory outcomes. Current Opinion in Microbiology. *7*, 102-108.

Bedinger, P., and Alberts, B.M. [1983]. The 3'-5' proofreading exonuclease of bacteriophage T4 DNA polymerase is stimulated by other T4 DNA replication proteins. J. Biol. Chem. *258*, 9649-9656

Boss, A., Nussbaum-Shochat, A., and Amster-Choder, O. [1999]. Characterisation of the Dimerisation Domain in BglG, an RNA-Binding Transcription antiterminator from *Escherichia coli*. J. Bacteriol. *181*, 1755-1766.

Brammar, W.J., Charles, I.G., Matfield, M., Liu, C.P., Drew, R.E., Clarke, P.H. [1987]. The nucleotide sequence of the *amiE* gene of *Pseudomonas aeruginosa*. FEBS Lett. *11*, 291-294.

Brammar, W.J., Clarke, P.H., and Skinner, A.J. [1967] Biochemical and genetic studies with regulator mutants of the *Pseudomonas aeruginosa* 8602 amidase system. J. Gen. Microbiol. *47*, 87-102.

Brammar, W.J., and Clarke, P.H. [1964]. Induction and Repression of *Pseudomonas aeruginosa* amidase. J. Gen. Microbiol. *37*, 307-319.

Brown, J.E., Brown PR., and Clarke, P.H. [1969]. Butyramide-utilizing mutants of *Pseudomonas aeruginosa* 8602 which produce an amidase with altered substrate specificity. J. Gen. Microbiol. *57*, 273-285.

Burgess, R., Travers, A. A., Dunn, J. J. and Bautz, E. K. [1969]. Factor stimulating transcription by RNA polymerase. Nature *221*, 43–46.

Butler, E.T., and Chamberlin, M.J. [1982]. Bacteriophage SP6-specific RNA polymerase. I. Isolation and characterization of the enzyme. J. Biol. Chem. *257*, 5772-5778.

Casarégola, S., Jacq, A., Laoudj, D., McGurk, G., Margarson, S., Tempête, M., Norris, V., and Holland, I.B. [1992]. Cloning and analysis of the entire *Escherichia coli ams* gene. *ams* is identical to hmp1 and encodes a 114 kDa protein that migrates as a 180 kDa protein. J. Mol. Biol. *228*, 30-40

Chai, W., and Stewart, V. [1998]. NasR, a novel RNA-binding Protein, Mediates Nitrate-responsive Transcription Antitermination of the *Klebsiella oxytoca* M5a1 *nasF* Operon Leader *in vitro*. J. Mol. Biol. *283*, 339-351.

Chai, W., and Stewart, V. [1999]. RNA Sequence Requirements for NasR-mediated, Nitrate-responsive Transcription Antitermination of the *Klebsiella oxytoca* M5a1 *nasF* Operon Leader. J. Mol. Biol. *292*, 203-216.

Clarke, P.H., Drew, R.E., Turberville, C., Brammar, W.J., Ambler, R.P., Auffret, A.D. [1981]. Alignment of cloned *amiE* gene of *Pseudomonas aeruginosa* with the N-terminal sequence of amidase. Biosci Rep. *1*, 299-307.

Correll, C.C., and Swinger, K. [2003]. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution.
RNA. *9*, 355-363.

Cossart, and Gicquel-Sanzey. [1985]. Regulation of expression of the crp gene of *Escherichia coli* K-12: *in vivo* study. J. Bacteriol. *161*, 454–457.

Costerton, J.W., Stewart, P.S., and Greenberg, E.P. [1999]. Bacterial Biofilms: A Common Cause of Persistent Infections. Science. *284*, 1318-1322.

Cousens, D.J., Clarke, P.H., and Drew, R.E. [1987]. The amidase regulatory gene (*amiR*) of *Pseudomonas aeruginosa*. J. Gen. Microbiol. *133*, 2041-2052.

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, and Kornberg R.D. [2000]. Architecture of RNA polymerase II and implications for the transcription mechanism. Science. *288*, 640-649.

Darst, S.A., Edwards, A.M., Kubalek, E.W., and Kornberg, R.D. [1991]. Three-dimensional structure of yeast RNA polymerase II at 16 A resolution. Cell. *66*, 121-128.

Day, M., Potts, J.R. and Clarke, P.H. [1975]. Location of genes for the utilisation of acetamide, histidine and proline on the chromosome of *Pseudomonas aeruginosa*. Genet. Res. Camb. *25*, 71-76.

Declerck, N., Vincent, F., Hoh, F., Aymerich S., and van Tilbeurgh H. [1999]. RNA Recognition by Transcription Antiterminators of the BglG/SacY Family: Functional and Structural Comparison of the CAT Domain from SacY and LicT. J. Mol. Biol. *294*, 389-402.

Declerck, N., Le Minh, N., Yang, Y., Bloch, V., and Aymerich, K. [2002]. RNA Recognition by Transcription Antiterminators of the BglG/SacY Family: Mapping of SacY RNA. J. Mol. Biol. *319*, 1035-1048.

DeLano, W.L. [2002]. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org

Doring, G., Taccetti, G., Campana, S., Festini, F., and Mascherini, M. [2006]. Eradiaction of *Pseudomonas aeruginosa* in cystic fibrosis patients. European Respiratory Journal. *27*, 653-658.

Drew, R.E., Clarke, P.H., and Brammar, W. [1980]. The construction *in vitro* of derivatives of bacteriophage lambda carrying the amidase genes of *Pseudomonas aeruginosa*. Mol Gen Genet. *177*, 311-320.

Drew, R.E., and Lowe, N. [1989]. Positive Control of *Pseudomonas aeruginosa* Amidase Synthesis is Mediated by a Transcription Aniti-Termination Mechanism. J Gen. Microbiol. *135*, 817-823.

Drew, R.E., and Haq, M. [2002. Chapter 26, Lessons from the *ami* operon. Pseudomonas: Virulence and Gene Regulation. Ramos, J.L. 425-443. Kluwer Academic Pub.

Dubendorff, J.W., and Studier, F.W. [1991]. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with *lac* repressor. J Mol Biol. *219*, 45-59.

Durham, D.R., and Phibbs, P.V. [1982]. Fractionation and characterization of the phosphoenolpyruvate: fructose 1-phosphotransferase system from *Pseudomonas aeruginosa*. J. Bacteriol. *149*, 534–541.

Farin, F., and Clarke, PH. [1978]. Positive regulation of amidase synthesis in *Pseudomonas aeruginosa*. J Bacteriol. *135*, 379-392.

Ferre-D'Amare, A. and Burley, S. [1997]. Dynamic Light Scattering in Evaluating Crystallizability of Macromolecules, Methods Enzymol. *276*, 157-166.

Fried, M.G., and Crothers, D.M. [1984]. Kinetics and mechanism in the reaction of gene regulatory proteins with DNA. J. Mol. Biol. *172*, 263-282.

Garner, M.M., and Revzin, A. [1981]. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res. *9*, 3047-3059.

Goldman, B.S., Lin, J.T., and Stewart, V. [1994]. Identification and structure of the *nasR* gene encoding a nitrate- and nitrite-responsive positive regulator of *nasFEDCBA* (nitrate assimilation) operon expression in *Klebsiella pneumoniae* M5al. J Bacteriol. *176*, 5077-5085.

Gourse, R.L., Ross, W., and Rutherford, S.T. [2006]. General Pathway for Turning on Promoters Transcribed by RNA Polymerases Containing Alternate σ Factors. J Bacteriol. *188*, 4589-4591.

Graille M, Zhou CZ, Receveur-Brechot V, Collinet B, Declerck N, and van Tilbeurgh H. [2005]. Activation of the LicT transcriptional antiterminator involves

a domain swing/lock mechanism provoking massive structural changes. J Biol Chem.*280*, 14780-14789.

Grainger DC, Aiba H, Hurd D, Browning DF, and Busby SJ. [2007]. Transcription factor distribution in *Escherichia coli*: studies with FNR protein. Nucleic Acids Res. *35*, 269-278.

Greenblatt, J., Mah, T.F., Legault, P., Mogridge, J., Li, J., and Kay, L.E. [1998]. Structure and mechanism in transcriptional antitermination by the bacteriophage lambda N protein. Cold Spring Harb Symp Quant Biol. *63*, 327-336.

Grodberg, J., and Dunn, J.J. [1988]. ompT encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. J Bacteriol. *170*, 1245-1253.

Gross, M., Skoczylas, B., and Turski, W. [1965]. Separation of Phenol and Ribonucleic Acid by Dextran Gel Filtration (Sephadex G-25). Anal. Biochem. *11*, 10-15.

Hanahan, D. [1983]. Studies on Transformation of *Escherichia coli* with Plasmids. J. Mol. Biol. *166*, 557-580.

Henkin, T., and Yanofsky, C. [2002]. Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. BioEssays. *24*, 700 -707.

Hermann, T. and Westhof, E. [1999]. Non-Watson-Crick base pairs in RNA-protein recognition. Chem. Biol. *6*, 335-343.

Herrero, M., De Lorenzo, V., and Timmis, K.N. [1990]. Transposon Vectors Containing Non-Antibiotic Resistance Selection Markers for Cloning and Stable Chromosomal Insertion of Foreign Genes in Gram Negative Bacteria. J Bacteriol. *172*, 6557-6567.

Hester, K.L., Lehman, J., Najar, F., Song, L., Roe, B.A., MacGregor, C.H., Hager, P.W., Phibbs, P.V. Jr, and Sokatch, J.R. [2000]. Crc is involved in catabolite repression control of the *bkd* operons of *Pseudomonas putida* and *Pseudomonas aeruginosa*. J Bacteriol. *182*, 1144-1149.

Heus, H.A., and Pardi, A. [1991]. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science. *253*, 191-194.

Hiratsu, K., Amemura, M., Nashimoto, H., Shinagawa, H., and Makino K. [1995]. The *rpoE* gene of *Escherichia coli*, which encodes sigma E, is essential for bacterial growth at high temperature. J. Bacteriol. *177*, 2918-2922.

Ingram, V.M. [1957]. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. Nature. *180*, 326-328.

Jacob, F., and Monod, J. [1961]. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. *3*, 318-3156.

Jaeger, J.A., Turner, D.H., and Zuker, M. [1989]. Improved predictions of secondary structures for RNA. Proc. Natl. Acad. Sci. USA. *86*, 7706-7710.

Kelly, M., and Clarke, P.H. [1962]. An Inducible Amidase Produced by a Strain of *Pseudomonas aeruginosa*. J. Gen. Microbiol. *27*, 305-316

Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. [2002]. Shortening of the RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. Mol. Cell, *10*, 1151-1162.

Kramer, A., Schwebke, I., and Kampf, G. [2006]. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. BMC Infectious Disease. *16*, 130-138.

Kuznedelov, K., Lamour, V., Patikoglou, G., Chlenov, M., Darst, S.A., and Severinov, K. [2006]. Recombinant *Thermus aquaticus* RNA Polymerase for Structural Studies. J. Mol. Biol. *359*, 110-121.

Kyte, J., and Doolittle, R. [1982]. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. *157,* 105-132.

Lejeune A, Vanhove M, Lamotte-Brasseur J, Pain RH, Frère JM, and Matagne A. [2001]. Quantitative analysis of the stabilization by substrate of *Staphylococcus aureus* PC1 beta-lactamase. Chem Biol. 8, 831-842

Li, J., Horwitz, R., McCracken, S., and Greenblatt, J. [1992]. NusG, a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage lambda. J. Biol. Chem. *267*, 6012-6019.

Lin, J.T., and Stewart, V. [1996]. Nitrate and nitrite-mediated transcription antitermination control of *nasF* (nitrate assimilation) operon expression in *Klebsiella pneumoniae* M5al. J Mol. Biol. *256*, 423-35.

Lowe, N., Rice, P.M., and Drew, R.E. [1989]. Nucleotide sequence of the aliphatic amidase regulator gene (*amiR*) of *Pseudomonas aeruginosa*. FEBS Lett. *27*, 39-43.

Maniatis, T., Fritsch, E.F., and Sambrook, J. [1989]. Molecular Cloning, A laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Manival, X., Yang, Y., Strub, M.P., Kochoyan, M., Steinmetz, M., and Aymerich, S. [1997]. From genetic to structural characterization of a new class of RNA-binding domain within the SacY/BglG family of antiterminator proteins. EMBO J. *16*, 5019-5029.

Marra, A.R., Bar, K., Bearman, G.M., Wenzel, R.P., and Edmond, M.B. [2006]. Systemic inflammatory response syndrome in nosocomial bloodstream infections with *Pseudomonas aeruginosa* and *Enterococcus* Species: comparison of elderly and nonelderly patients. J. Am. Geri. Soc. *54*, 804-810.

Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. [1999]. Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. J. Mol. Biol. *288*, 911-940.

McClure, W.R. [1985]. Mechanism and control of transcription initiation in prokaryotes. Ann. Rev. Biochem. *54*, 171-204.

Meka, H., Werner, F., Cordell, S.C., Onesti, S., and Brick, P. [2005]. Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. Nucleic Acids Res. *33*, 6435–6444.

Miller, V.L., and Mekalanos, J.J. [1988]. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholorae* requires toxR. J. Bacteriol. *170*, 2575-2583.

Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C., [1987]. Oligonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res. *15*, 8783-8798.

Moffat, B.A., and Studier, F.W. [1986]. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. *189*, 113-130.

Morth, J.P., Feng, V., Perry, L.J., Svergun, D.I., and Tucker, P.A. [2004]. The crystal and solution structure of a putative transcriptional antiterminator from *Mycobacterium tuberculosis*. Structure. *12*, 1595-1605.

Murakami, K.S, Masuda, S., and Darst, S.A. [2002]. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. Science. *296*, 1280-1284.

Norman, R.A., Laa Poh, C., Pearl, L., O'Hara, B.P., and Drew R.E. [2000]. Steric Hindrance Regulation of the *Pseudomonas aeruginosa* Amidase Operon. The Journal of Biochemical Chemistry. *275*, 30660-30667.

Novy, R., Yaeger, K., Held, D. and Mierendorf, R. [2002]. Co-expression of multiple target proteins in *E. coli*. In*Nova*tions. *15*, 2-6

O'Hara, B.P., Norman, R.A., Wan, P.T.C., Roe, S.M., Barrett, T.E., Drew, R.E., and Pearl, L. [1999]. Crystal structure and induction mechanism of AmiC-AmiR: a ligand-regulated transcription antitermination complex. EMBO J. *18*, 5175-5186.

O'Hara, B.P., Wilson, S.A., Lee, A.W., Roe, SM, Siligardi, G., Drew, R.E., and Pearl, L.H. [2000]. Structural adaptation to selective pressure for altered ligand

specificity in the *Pseudomonas aeruginosa* amide receptor, *amiC*. Protein Eng.*13*,129-132.

Ouhammouch, M., Werner, F., Weinzierl, R., and Geiduschek, E.P. [2004]. A fully recombinant system for activator-dependent archaeal transcription. J. Biol. Chem. *279*, 51719-51721

Oxender, D.L., Anderson, J.J., Daniels, C.J., Landick, R., Gunsalus, R.P., Zurawski, G., Selker, E., and Yanofsky, C. [1980]. Structural and functional analysis of cloned DNA containing genes responsible for branched-chain amino acid transport in *Escherichia coli*. Proc Natl Acad Sci U S A. *77*, 1412-1416.

Palleroni, N.J. [1992]. Present situation of the taxonomy of aerobic Pseudomonads. *Pseudomonas*: Molecular Biology and Biotechnology. American Society for Microbiology. 105-115.

Parkinson, J.S. [1993]. Signal Transduction Schemes of Bacteria. Cell. *73*, 857-871.

Pearl, L., O'Hara, B., Drew, R.E., and Wilson, S.A. [1994]. Crystal structure of AmiC: the controller of transcription antitermination in the amidase operon of *Pseudomonas aeruginosa*. EMBO J. *13*, 5810–5817.

Peritz, A.E., Kierzek, R., Sugimoto, N., and Turner, D.H. [1991]. Thermodynamic study of internal loops in oligoribonucleotides: symmetric loops are more stable than asymmetric loops. Biochemistry. *30*, 6428-6436.

Phillips, A.T., and Mulfinger, L.M. [1981]. Cyclic Adenosine 3',5'–Monophosphate Levels in *Pseudomonas putida* and *Pseudomonas*

aeruginosa During Induction and Carbon Catabolite Repression of Histidine Synthesis. J. Bacteriol. *145*, 1286-1292.

Plapp, B.V. [1995]. Site-directed mutagenesis: a tool for studying enzyme catalysis. Methods Enzymol. *249*, 91-119.

Polyakov, A., Severinova, E., and Darst, S. [1995]. Three-Dimensional Structure of *E. coli* Core RNA Polymerase: Promoter Binding and Elongation Conformations of the Enzyme. Cell. *83*, 365-373.

Rao, S.T., and Rossmann, M.G. [1973]. Comparison of super-secondary structures in proteins. J. Mol. Biol. *76*, 241-256.

Scheler, C., Lamer, S., Pan, Z., Li, XP., Salnikow, J., and Jungblut, P. [1998]. Peptide mass fingerprint sequence coverage from differently stained proteins on two-dimensional electrophoresis patterns by matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS). Electrophoresis. *19*, 918-927.

Schilling, O., Langbein, I., Muller, M., Schmalisch, MH., and Stulke, J. [2004]. A protein-dependent riboswitch controlling *ptsGHI* operon expression in *Bacillus subtilis*: RNA structure rather than sequence provides interaction specificity. Nucleic Acids Res. *20*, 2853-2864.

Schilling, O., Herzberg, C., Hertrich, T., Vorsmann, H., Jessen, D., Hubner, S., Titgemeyer, F., and Stulke, J. [2006]. Keeping signals straight in transcription regulation: specificity determinants for the interaction of a family of conserved bacterial RNA-protein couples. Nucleic Acids Res. *34*, 6102-6115.

Schnetz, K., Stulke, J., Gertz, S., Kruger, S., Kreig, M., Hecker, M., and Rak, B. [1996]. LicT, a *Bacillus subtillis* transcriptional antiterminator protein of the BglG family. J. Bacteriol. *178*, 1971-1979.

Shafikhani, S., Siegel, R.A., Ferrari, E., and Schellenberger, V. [1997]. Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. Biotechniques. *23*, 304-310.

Shu, C.J., and Zhulin, I. [2002]. ANTAR: an RNA-binding domain in transcription antitermination regulatory proteins. TRENDS in biochemical Sciences. *27*, 3-5.

Skordalakes, E., and Berger, J.M. [2003]. Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. Cell. *114*, 135-146.

Smith, D.B., and Johnson, K.S. [1988]. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene. *67*, 31-40.

Stevens, A. [1960]. Incorporation of the adenine ribonucleotide into RNA by cell fractions from *E. coli* . Biochem. Biophys. Res. Commun. *3*, 92–96.

Stover, C.K., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K.S., Wu, Z., Paulsen, I., Reizer, J., Saier, M.H., Hancock, R.E., Lory, S., and Olson, M.V. [2000]. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature, *406*, 959-964.

Studier, F.W., and Rosenburg, A.H. [1981]. Genetic and Physical Mapping of the late region of bacteriophage T7 DNA by use of cloned fragments of T7 DNA. J. Mol. Biol. *153*, 503-525.

Studier, F.W., and Moffatt, B.A. [1986]. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol. *189*, 113-130.

Stulke, J., Arnaud M., Rapoport G., and Martin-Verstraete, I. [1998]. PRD--a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. Mol Microbiol. *28*, 865-74.

Tabor, S., and Richardson, C.C. [1985]. A bacteriphage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. U S A. *82*, 1074-1078.

Turner, D.H., Sugimoto, N., and Freier, S.M. 1988. RNA structure prediction. Ann. Rev. Biophys. Biophys. Chem. *17*, 167-192.

Vandeyar, M.A., Weiner, M.P., Hutton, C.J., and Batt, C.A. [1988]. A simple and rapid method for the selection of oligodeoxynucleotide-directed mutants. Gene. *65*, 129-133.

Wade, J.T., Roa, D.C., Grainger, D.C., Hurd, D., Busby, S.J.W., Struhl, K., and Nudler, E. [2006]. Extensive functional overlap between σ factors in *Escherichia coli*. Nature Structural and Molecular Biology. *13*, 806-814.

Walker, P.A., Leong, L.E., Ng, P.W., Tan, S.H., Waller, S., Murphy, D., and Porter, A.G. [1994]. Efficient and rapid affinity purification of proteins using recombinant fusion proteases. Biotechnology. *12*, 601-605.

Wang, D., Severinov, K., and Landick, R. [1997]. Preferential interaction of the *his* pause RNA hairpin with RNA polymerase *beta* subunit residues 904-950 correlates with strong transcriptional pausing. Proc Natl Acad Sci U S A. *94*, 8433-8438.

Wang, Y., and DeHaseth, P.L. [2003]. Sigma 32-dependent promoter activity in vivo: sequence determinants of the *groE* promoter. J. Bacteriol. *185*, 5800-5806.

Weiss, S., and Gladstone, L. [1959]. A mammalian system for the incorporation of cytidine triphosphate into ribonucleic acid. J. Am. Chem. Soc. *81*, 4118–4119.

Wigneshweraraj, S.R., Burrows, P.C., Severinov, K., and Buck, M. [2005]. Stable DNA opening within open promoter complexes is mediated by the RNA polymerase β'-Jaw domain. J. Biol. Chem. *280*, 36176-36184.

Wilson, S.A., and Drew, R.E. [1991]. Cloning and DNA sequence of amiC, a new gene regulating expression of the *Pseudomonas aeruginosa* aliphatic amidase, and purification of the *amiC* product. Journal of Bacteriology. *173*, 4914-4921.

Wilson, S.A., Wachira, S.J., Drew, R.E., Jones, D., and Pearl, LH. [1993]. Antitermination of amidase expression in *Pseudomonas aeruginosa* is controlled by a novel cytoplasmic amide-binding protein. EMBO J. *12*, 3637–3642.

Wilson, S.A., Williams, R.J., Pearl, L., and Drew, R.E. [1995]. Identification of Two New Genes in the *Pseudomonas aeruginosa* Amidase Operon, Encoding an ATPase [AmiB and a Putative Intergral Membrane Protein [AmiS]. J. Biol. Chem. *270*, 18818-18824.

Wilson, S.A., Wachira, S., Norman, R.A., Pearl, L., and Drew, R.E. [1996]. Transcription antitermination regulation of the *Pseudomonas aeruginosa* amidase operon. EMBO J. *15*, 5907-5916.

Yang, Y., Declerck, N., Manival, X., Aymerich, S., and Kochoyan, M. [2002]. Solution structure of thr LicT-RNA antitermination complex: CAT clamping RAT. EMBO J. *21* 1987-1997.

Yanisch-Perron., C., Vieira, J., and Messing, J. [1985]. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene. *33*, 103-119.

Yarnell, W.S., and Roberts, J.W. [1999]. Mechanism of Intrinsic Transcription Termination and Antitermination. Science. *284*, 611-615.

Ye, J.J., and Saier, M.H., Jr. [1996]. Regulation of sugar uptake via the phosphoenolpyruvate-dependent phosphotransferase systems in *Bacillus subtilis* and *Lactococcus lactis* is mediated by ATP-dependent phosphorylation of seryl residue 46 in HPr. J Bacteriol. *178*, 3557-3563.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. [1999]. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. Cell. *17*, 687-690.

Zinkel, S.S., and Crothers, D.M. [1991]. Catabolite activator protein-induced DNA bending in transcription initiation. J. Mol Biol. *20*, 201-215.

Zubay, G., Schwartz, D. & Beckwith, J. [1970]. Mechanism of activation of catabolite-sensitive genes: a positive control system. Proc. Natl Acad. Sci. USA *66*, 104–111.

Zuker, M., Jaeger, J.A., and Turner, D.H. [1991]. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. Nucleic Acids Res. *19*, 2707-2714.

Zuker, M. [2003]. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *13*, 3406-3415.